How to help restaurants survive COVID-19

Project Scoping

Overview of industry, business, or problem

Congress passed a $25 billion COVID-19 bailout for the airline industry but not one tailored to the restaurant industry, which is four times bigger in terms of sales and 18 times bigger in number of jobs (restaurant industry is the nation's second-largest private-sector employer with an employee base of 15.6 million). Restaurant industry losses are on track to top $240 billion by the end of 2020 — more than any other industry.

Define the specific problem that should be solved

How to help the restaurant industry survive?

- Identify who needs a delivery partner, switch to curbside pick-up or keep their strategies.

How to answer those questions:

- Analysis of consumer preferences (availability to coming back to restaurants or not, likely to buy food through delivery services, curbside pick-up or in-store, likely to use a mobile app to buy food, contactless payment methods or cash instead).
- Analysis of foot traffic and mobility patterns: detecting businesses in areas with less foot traffic, businesses far away from recreational areas, transit stations, in areas where people are less likely to come back to restaurants and are more concerned about getting the virus.

Why does this problem matter?

Restaurants need to know what investments are likely to help them survive. With consumer fears over human-to-human contact at an all-time high, tools that allow restaurants to conduct business while eliminating touchpoints have risen in popularity. **This trend is likely to continue even once dining rooms reopen**, so restaurants will be investing in systems that support contactless dinings, such as mobile payment and ordering.

Regardless of when restaurants decide to reopen, **the takeout orders** that allow them to hang on during the roughest weeks **will continue to be essential to the slow rebuilding of their business during the next year**.

Potential Audience

- Tech companies should be one of the most interested audiences, in the sense that restaurants are going to need to invest in more low contact technologies to provide contactless dinings (mobile payment, ordering). That translates into apps and websites usable by all age groups, investment in network security, tools to make delivery and pick-up smoother, apps to improve the pick-up experience and also some kind of share apps where diners can place a single order and pick up multiple items at different brands.
- Community Supported Agriculture, Community Farmers Markets: Their partnerships with restaurants are crucial to keep their own operations.

Dataset Details

Table 1 summarizes the data sources used for the project, all of them publicly available, under the following categories:

- Bureau of Economic Analysis Data (GDP, Personal Consumption, Income, and Employment)
- Federal Reserve Bank of St. Louis (Unemployment)
- Annual Retail Trade Survey (Monthly Retail Sales and Inventories)
- Mobility Patterns (Apple Mobility Reports, Descartes Lab Mobility Changes, Google Community, Foursquare Community Mobility Data)
- Household Pulse Survey 2020
- Restaurants platforms (Yelp dataset, OpenTable Data)

Every category includes a list of datasets, main features, and topics of interest.

Category	Datasets	Source	Features	Topics	Related topics
Bureau of Economic Analysis Data	1. Domestic Product and Income by Industry and Expanded Detail 2. Personal Consumption Expenditures by Major Type of Product 3. Income and Employment by Industry (until 2019)	https://www.bea.gov/data/by-place-us	Measured per qtr; will be updated Nov 25.	Trends of GDP in food service, employment, incomes(Table 1.5.3. Real Gross Domestic Product, Expanded Detail, Quantity Indexes)	Determine GDP trends in food services vs food purchased during the pandemic. Compare performance with other categories of durable-nondurable goods and services.
				Analysis of expenditures in food purchased and food services (Table 1.5.1. Percent Change From Preceding Period in Real Gross Domestic Product, Expanded Detail and Table 1.5.2. Contributions to Percent Change in Real Gross	Determine the change in personal consumption expenditures in food purchased for consumption vs food services per qrt 2018-2020.

				Domestic Product, Expanded Detail)	
Federal Reserve Bank of St. Louis	Monthly Unemployment Numbers by Industry	https://fred.stlouisfed.org/release/tables?rid=50&eid=4635#snid=4770 https://drive.google.com/drive/folders/14cO5A5K0ulc16antLYLwJEjv95So1YLM	Measured Monthly and separated by Industry. Includes both percentages as well as raw numbers Includes monthly data from 2020 alone along with data from 2005 to 2020.	Trends of Unemployment and how it affected each industry.	Determining how this recession compares to that of 2008 and how long it took for things to recover. Month to Month change by industry can determine which industries are recovering and if their recovery is influencing the restaurant industry. Analysis of whether these trends are shown in our other graphs.
Monthly Retail and Food Services Sales and Inventories (Annual Retail Trade Survey)	1. Estimates of Monthly Retail and Food Services Sales 2. Estimates of Monthly Retail Inventories/Sales Ratios	https://drive.google.c fromom/drive/folders/1HjBamjglzwTK_PZ7VxyPBZDtPe7nWiTk	Data 1992 to 2020. 3 Datasets: First on sales, second on inventory/sale percentage, and third on current data not incorporated into the first dataset. Measured monthly with a cumulative total for each year up to 2019. 2020 is also measured monthly but with a predicted annual total. Each sheet in the excel database contains both adjusted and non-adjusted sales. 2 Dataframes: First - Shape(28x16) contains annual sales totals in relevant industries. Second - Shape(336x16) contains monthly sales in relevant industries.	Compare performance of food/beverage stores vs food services for a duration of three decades. Example entries: Retail sales, total, Restaurants and other eating places, Full-service restaurants	Trend Analysis: Industries affected by covid. Factors: 2008 Recession, DotCom Bust, COVID-19, Seasons, inflation. Graphs showing the evolution of sales in food services over years and on a monthly basis.
Mobility Patterns	1. Apple Mobility Reports	https://drive.google.com/drive/folders/1NB1oEsFE33XczpebAINhuFH828U2fCbk	The relative amount of route requests from every region/date. Information broken into states, counties, date from Jan 2020-October 2020. 3 datasets with information by state and 3 datasets (complete) with information by county (incomplete). States datasets by type of transportation (transit, driving, walking). 50 rows	Transit	Can look at transportation type preferences by county: walking, driving, transit. Identify zones with more foot traffic.

			(states), 293 variables, 5 categorical variables and 288 numerical variables representing number of route requests for each state from January to October 2020.		
	2. Descartes Lab Mobility Change	https://drive.google.com/drive/folders/1oDDY1Vhpaxa29Iy1_Mj6zU1aZw6pEkGr	The distance a typical member of a given population moves in a day (kms). 2 types of datasets: **The median of the max-distance mobility for all samples(m50)** in the specified region and **the percent of normal m50 in the region, with normal m50 defined during 2020-02-17 to 2020-03-07 (m50_index).** Information broken into states, counties, dates. Number of datasets with county divisions: 2 and number of datasets with state division: 2 Datasets with county divisions: 3k rows, 247 variables. 4 categorical variables, 242 numerical m50/m50_index. Dates from 03-01-2020 to 10-30-2020. Datasets with states divisions: 51 rows, 247 variables. 3 categorical variables, 242 numerical m50/m50_index. Dates from 03-01-2020 to 10-30-2020.	Transit	Compare mobility before and after lockdowns in the different states for Milestone 1. Once we choose a state or city, use the mobility by county to identify the areas more affected by lockdowns (Milestone 2).
	3. Google Community Mobility Data	https://www.google.com/covid19/mobility/	Global information is broken down into counties. Keeps track of mobility changes as a percent. Industries tracked (6 different types of places): grocery and pharmacy, parks, parks, transit stations, retail and recreation,	Transit	Can look at public transport changes, mobility for different purposes. The trends of mobility to groceries, parks, workplaces and residential give us indirect information of the potential flow of people to restaurants close to those areas.

			workplaces. 2 types of datasets: mobility by county and state. 51 rows, 258 columns, 2 categorical variables, 256 numerical variables representing visits and duration of visits to different places between February and October 2020 compared to a baseline.		
	4. Foursquare Community Mobility Data	https://drive.google.com/drive/folders/1_lJWAae0MtYr7U_k3RUqBcaDhqa3siRv	Visits, average duration in minutes and median visit length in minutes to 25 categories of places. Based on 13 million users from 01-01-2020 to 10-29-2020 by state. 52 rows, 302 columns, 2 categorical variables and 303 numerical variables representing the number of visits, average duration of visits and median duration of visits to Food stores and Food restaurants by state.	Visits, duration of visits to Food stores and Fast Food Restaurant.	Compare visits to food stores and fast food by states before and after the lockdowns (data available from January).
	5. Foursquare COVID-19 National and Regional	https://console.aws.amazon.com/dataexchange/home?region=us-east-1#/subscriptions/prod-hwaqvsrhtj7hm	AWS Data Exchange. Indexed foot traffic to 19 categories of venues. The indexed data is broken out geographically, with included data for National, SF, NYC, LA, and Seattle. The data is normalized against U.S. Census data to remove age, gender and geographical bias. Data is provided daily from 02/19/2020.	Updated daily foot traffic information splitting dining in casual and fast food restaurants (national level) and by city.	Compare visits to food stores and fast food by states during the entire year.
Household Pulse Survey		https://www.census.gov/data/experimental-data-products/household-pulse-survey.html	17 weeks from April 23 to October 26 2020. Surveyed people between 50k-100k per week. Variables between 82 to 188, mostly categorical.	Affordability of food, free meals and spending use of the Economic Impact Payment	Recognize groups eligible for the social food programs but not included (insights for the National Association of Restaurants).
				Shopping and purchase preferences.	Shopping modalities, payment modalities, resumed/avoided eating at restaurants. Use of credit

			Missing data designed as -88 and -99. Require use of a data dictionary to translate the name of columns and categories. Demographics, spending, food, shopping, teleworking, trip trends variables. Dataset includes sub variables (secondary questions of the survey) which values depend on the answers to primary questions. In consequence, there are missing values in all the secondary variables and they will be removed during their specific analysis.		cards, apps to buy online. Consumer preferences (prepared food vs ingredients to cook at home)
				Trips and teleworking variables	Fewer transit trips, planned trips, trips to stores (give us information about likely to leave the home to buy meals vs use of delivery)
Restaurants	Yelp dataset	https://drive.google.com/drive/folders/1mp2texeym4VJbnPQFFFxnYyNFMjNRInu	Name, location (state and county), status (open, closed), attributes (take-out, outdoor dining, parking), categories (type of food), hours, stars, reviews.	Restaurant current status current services offered, location and popularity	Can determine how the restaurants were faring pre pandemic. Ratings and review count give us clues into how popular/competent these places may have been. We could also potentially find out more into how these restaurants responded to Covid (hours, takeout options)
	OpenTable dataset	https://drive.google.com/drive/folders/1BspmA9jUOuXjVOrTBeGa8-h5DSiZP3On	Sample of +20k restaurants across the country in the OpenTable network (online reservations, phone reservations and walks-in). States and metros with +50 restaurants on the OpenTable platform.	Tracking seated diners related to the same dates in 2019. Do not require a cleaning process.	Overall impact of COVID-19 in the industry showing year over year seated diners at a sample of restaurants.

Table 1: Summarize data sources

Data Wrangling

Data Cleaning

Table 2 explores the data cleaning steps required in every dataset and how the methodology used assures that the Data is ready for the Exploratory Data Analysis. Data acquisition includes direct download of Excel, CSV and json files from the corresponding websites and web-scraping. Data cleaning incorporates the use of regular expressions, missing values exploratory methods consolidated in a Python script, extraction of the variables of interest per dataset, exploration of particular inconsistencies and development of specific methods according to the nature of the dataset.

Data is mostly numerical (GDP, Personal Consumption Expenditures, Unemployment, Sales, Inventories, Seated dining at restaurants, changes in mobility patterns, number of visits to places, duration of visits) except by the Household Pulse Survey (variables are categories representing answers to the survey) and Yelp Dataset (include categories of food, services, location).

Dataset	Cleaning steps	Why it is required?
Household Pulse Survey	Build sub-dataset of spending: extract EIP and EIPSPND variables over weeks and demographics	Expenditure patterns: Track percent change of people receiving EIP over weeks and its use by demographics
	Build sub-dataset of shopping variables over weeks and demographics: 1. extract CHNGHOW and WHYCHNGD variables. 2. extract FEWRTRIPS, FEWRTRANS variables 3. extract EXPNS_DIF: difficulty with expenses	1. Changes in shopping: purchases modalities, cash/credit card, avoid/resume dining in restaurants and reason. Track percent change over weeks and group by demographics. 2. transit trips and trips to stores: identify groups less likely to leave their homes 3. Relation between EIP and EXPNS_DIF
	Food Sufficiency over weeks and demographics 1. extract FOODSUFRSN (food sufficiency), FREEFOOD, WHEREFREE (free groceries), SNAP_YN, PRIFOODSUF.	The NRA is asking to expand the eligibility to RMP as part of SNAP. EDA related to the use of SNAP and restaurants struggles by state. How many people receive SNAP benefits? People that can't get out to buy or they are afraid. How many delivery services the city needs?
	Methods: 1. Incorporate age of the surveyed people, replace codes with nan values, drop duplicates in weekly analysis and include the dates of the survey. 2. Identification of missing values over rows and columns	Age instead birthday year for age groups analysis. Deal with NaN instead of numerical codes for null responses. Avoid duplicates surveyed people present in more than one week and use of dates for better reference.
Descartes Lab Mobility Change (Traffic)	1. Identification of missing values over rows and columns. Drop nan rows and columns in county and states datasets. 2. Extract STATE, COUNTY and m50/m50 index from 03-01-2020 to 10-30-2020. Drop the rest of the columns.	Data is going to be pivoted to visualize trends over the year by state. We'll identify states more affected for the lockdowns using m50_index and compare general trends using m50.
Foursquare Community Mobility Data (Traffic)	1. Identification of missing values over rows and columns. Drop nan rows (extra-row without values was removed). 2. Concatenation of the 6 datasets related to Food and Fast Food mobility.	Analysis of mobility related to Food stores and Food restaurants. Compare these results with m50 data.

Apple Mobility Reports (Traffic)	1. Identification of missing values over rows and columns. 4 states have missing values. We are going to fill them with the median of the rest of the states instead of dropping them. 2. Concatenation of the 3 datasets related to type of transportation.	Analysis of mobility related to type of transportation by state before and after the lockdowns. Compare these results with m50 data and foursquare trends.
Google Community Mobility Data	1.Concatenation of categories of places to generate one dataset by states and another by counties. 2. Identification of missing values over rows and columns over the two datasets. Drop unuseful columns but keep the gaps(missing values) in mobility by counties.	The dataset by states is going to be used to complete the analysis of mobility by state. The dataset by counties will be used in the Milestone 3, when we need more information by counties looking for restaurants near parks, groceries, transit stations, residential and workplaces.
Yelp Dataset	1. Take the 500k rows from businesses dataset and select ones categorized "restaurants" using string match. 2. Clean the reviews dataset by replacing the missing values 3. Use geographic data to find the state and county for each business in the business dataframe 4. Calculate review count and average stars using the reviews dataframe for each business in business dataframe 5. Look at unique values to get a sense of what our data is and what problems we may run into	1. We need to seperate restaurants from everything else 2. We need clean data to accurately sift through our findings 3. Need to find county and state to get a better idea on our trends based on location 4. Gives us a good idea on how said restaurants are faring or fared in terms of popularity and competency 5. We want to see if there are any issues with our data for later versions
Domestic Product and Income by Industry and Expanded Detail (Table 12)	1. Remove all data under quarters 2016 - 2017 as it's blank and not relevant. 2. Align data points according to universal columns for quarter and summarized year (q1 of 2018 - q3 of 2020)	Analysis of service of foods (i.e. restaurants and bars) vs other categories as part of GDP. Assessing trends q3 2016 - q2 of 2020. Seeing relationship to fishing and farming commodity impact to determine relationship.
Personal Consumption Expenditures by Major Type of Product (Tables 1.5.2, 1.5.1, 1.5.3)	1. Remove all data under quarters 2016 - 2017 as it's blank and not relevant. 2. Align data points according to universal columns for quarter and summarized year (q1 of 2018 - q3 of 2020) 3. Identification of inconsistent rows, relative to the other data sets, which are not relevant to the subject matter. Drop rows: [Percent change at annual rate:] and [Percentage points at annual rates:]	See where personal consumption and capital has been spent over time (in particular before and after the impact of COVID. Correlate quarter results with that of GDP and income per industry. Review impact of imports vs exports as it pertains to business and service over same time series.
Income and Employment by Industry (until 2019)	1. Remove all data under quarters 2016 - 2017 as it's blank and not relevant. 2. Align data points according to universal columns for quarter and summarized year (q1 of 2018 - q3 of 2020)	This encapsulates the impact of COVID as it pertains to employment. Marrying income and employment numbers over the same time series with GDP determines how one affects the other. It can also determine the overall health (or risk) of the food services industry (in relation to others) and show impact for the need of relief for workers.
Unemployment by Month (Federal Reserve Bank of St Louis)	1. Import the data from source and download as an excel sheet 2. Replace all the column names with the name of the category they represent 3. Isolate the year and categories we want (Total Unemployment as well as Leisure and Hospitality for 2020)	The dataset was very clean to begin with but the column names had to be changed as before it was just ID numbers. Changing the names to what the column represented is necessary for readability. For Version One of the project we only need Leisure and Hospitality as well as total unemployment for 2020
Monthly Retail and Food Services Sales 1992-2020 (Annual Trade Survey)	1. Prep excel docs by formatting columns and rows for quick read_excel call by python. This included removing the irrelevant NAICS code, comments and white space. 2. Combine all sheets in the Excel data spread into one dataframe. Each sheet contains data for a year of sales in the various industries. 3. While combining, only add to dataframe relevant industries. Relevance is determined by proximity to the restaurant industry and thus includes grocery sales, etc. 4. Separate df into two new dataframes, one including annual sales, the other including monthly sales.	Most values were already clean as this is a public, comercial dataset. However there was a lot of superfluous information as regards the scope of our project which needed to be cut out. Making two dataframes, one of monthly and the other annually, allows for better analysis of data from different perspectives. Transposing time to rows allows for better plotting of line graphs.

	5. Drop columns containing all NaN values, otherwise ignore. This mostly was for the non-adjusted sales. 6. Reset Indices and transpose dfs to make the Kind of Business the key index and the columns be time.	
Foursquare + Apptopia	1. Join Datasets with aggregate information by type of food. 2. Join Datasets with information by individual dinings 3. Analysis of missing data	Year over year analysis of restaurant app usage vs foot traffic to analyze correlation between app usage and foot traffic in different dinings and types of food.

Table 2: Data cleaning process by dataset

The following diagrams show the methods applied in every dataset.

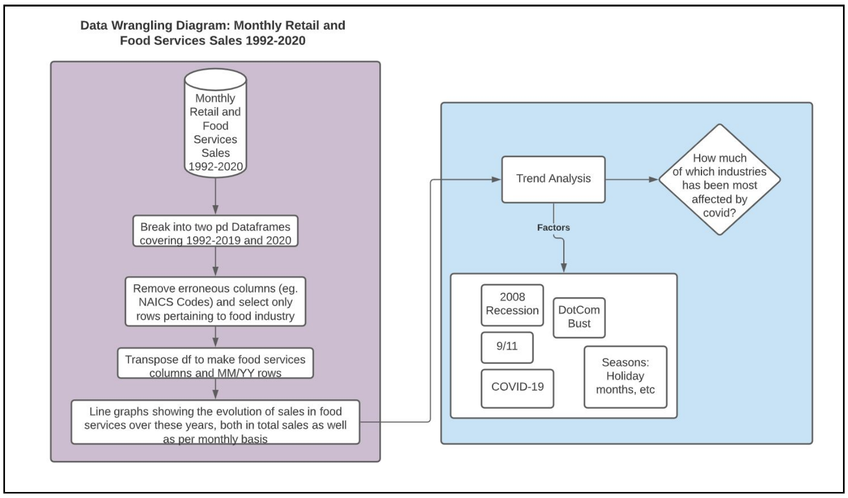


Diagram 1: Data Cleaning Process of Annual Trade Survey

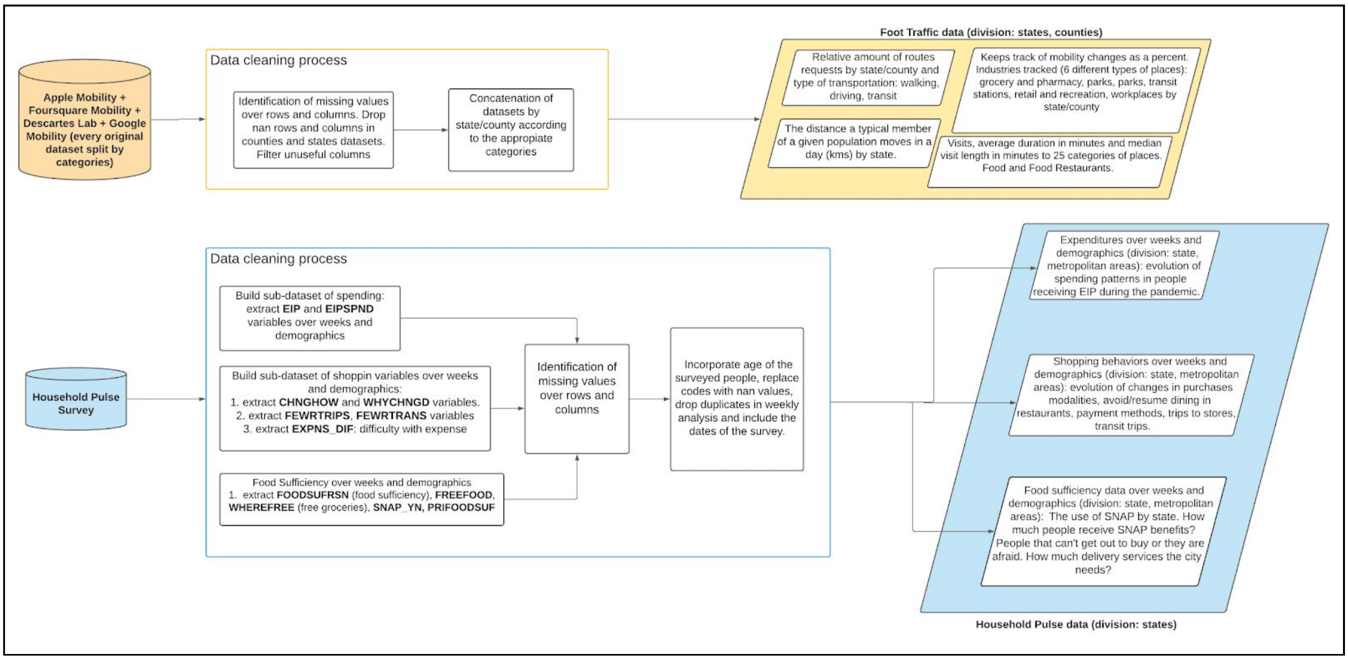


Diagram 2: Data Cleaning Process of Household Survey and Mobility Datasets

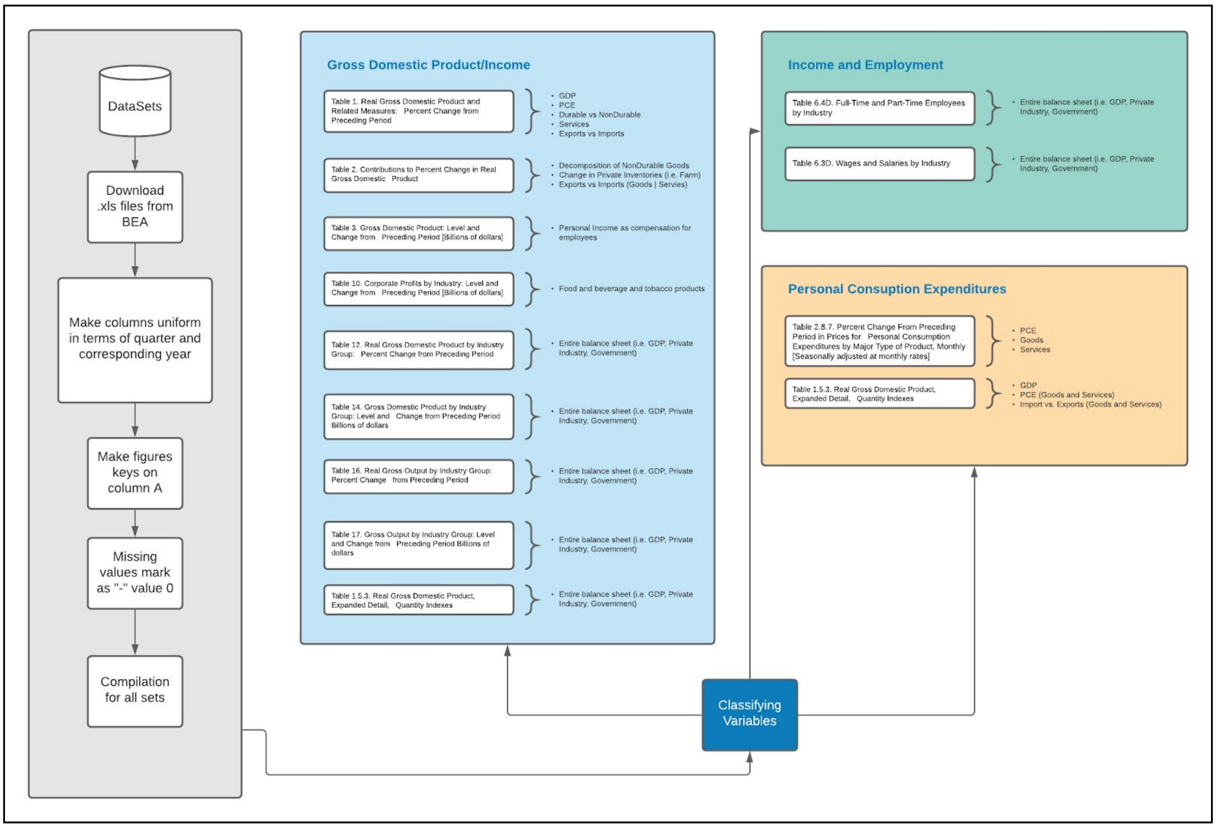


Diagram 3: Data Cleaning Process of BEA Datasets

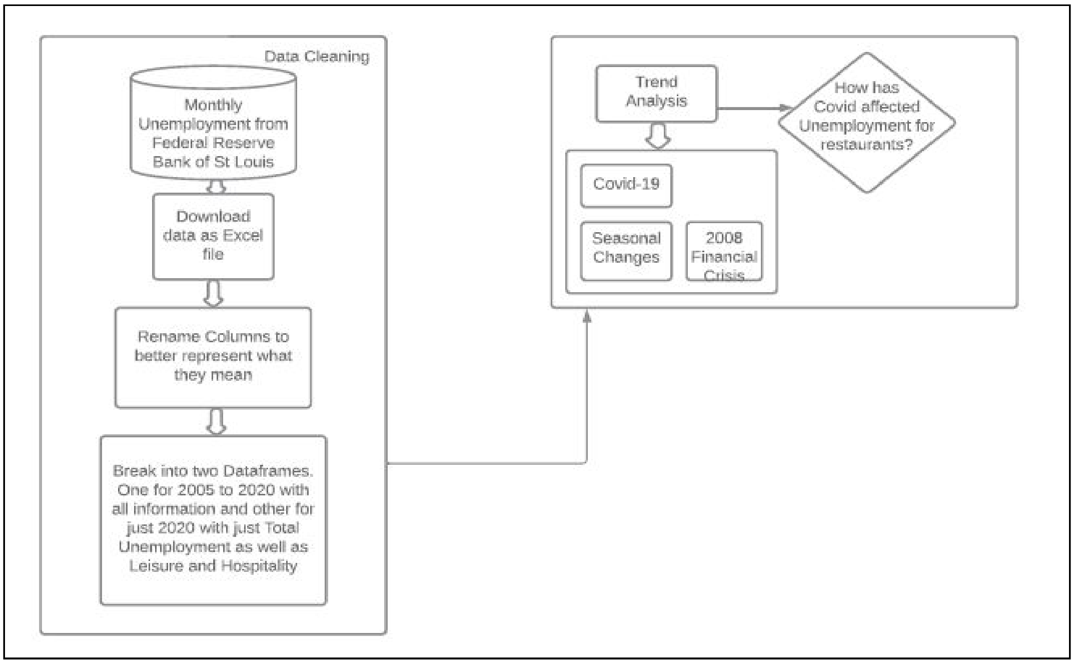


Diagram 4: Data Cleaning Process of Unemployment

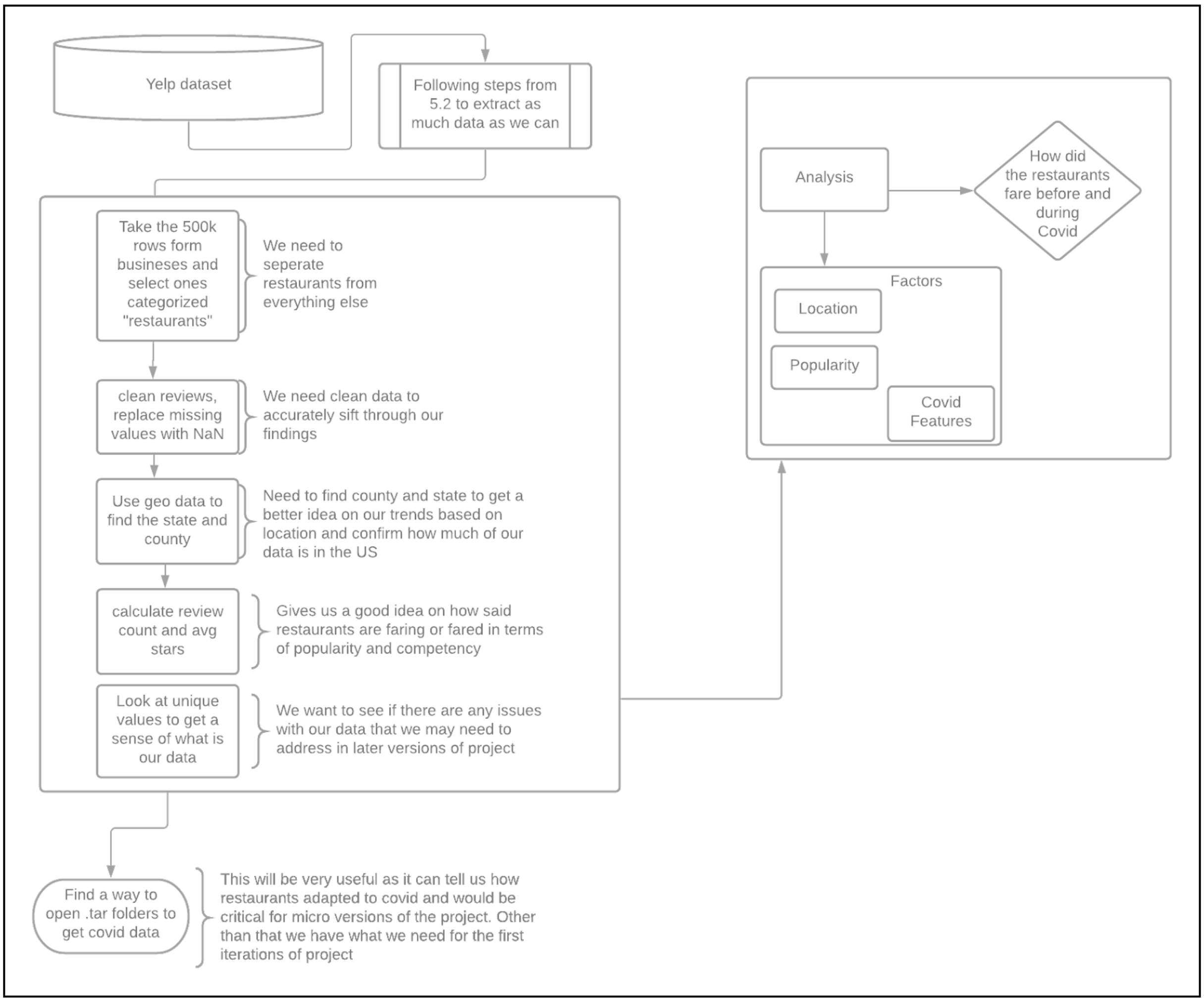


Diagram 5: Data Cleaning Process of Yelp dataset

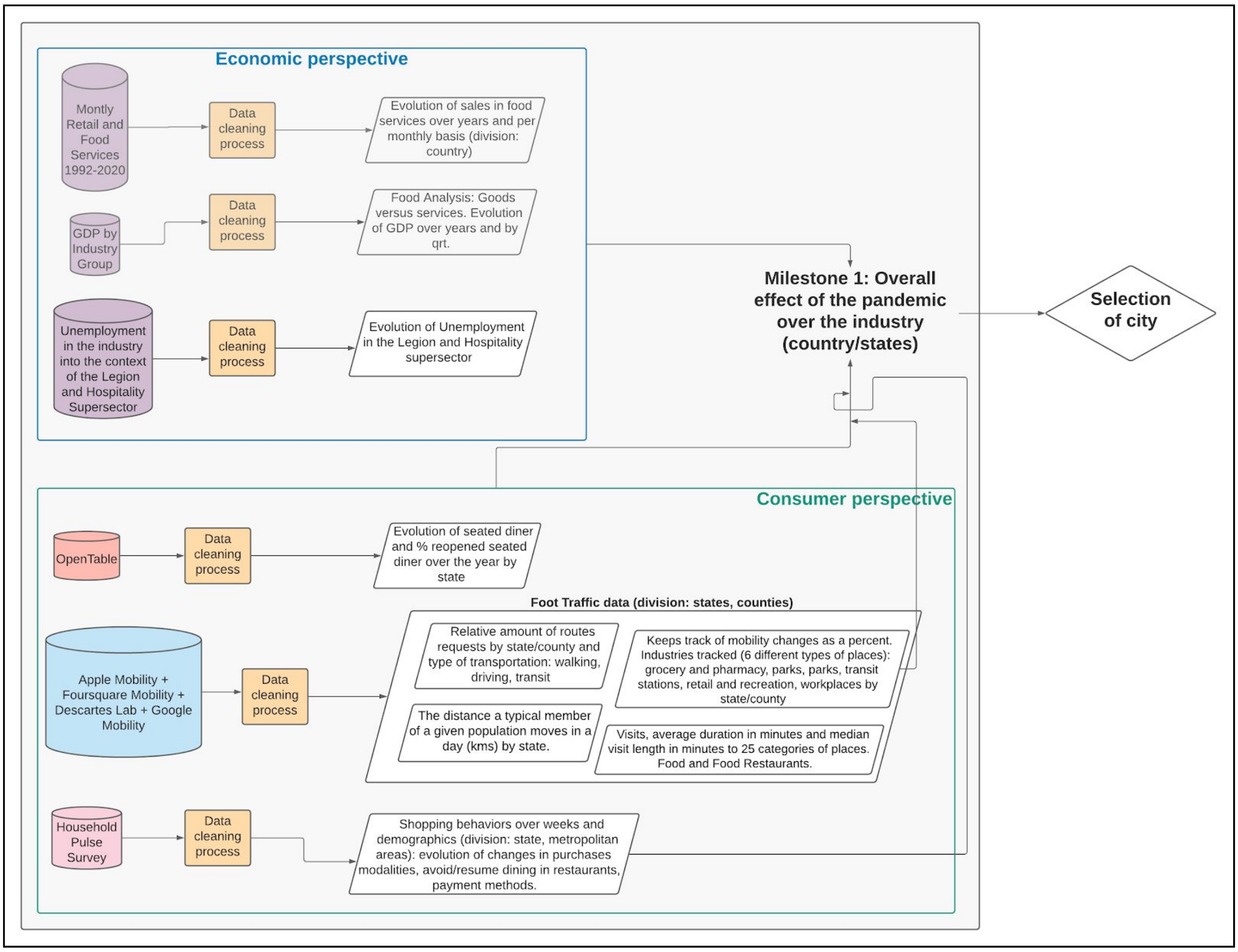


Diagram 6: Structuring Datasets for EDA Milestone 1

Data Structuring

The diagram 6 explores how datasets are going to be used to achieve the Milestone 1, which is an overall effect of the pandemic over the industry during this year from an economic perspective (tracking GDP variables, sales and inventories from the food industry as goods and services and employment variables) and a consumer perspective, looking for changes in consumer behaviours, mobility patterns related to the acquisition of food (as good and services). This macro analysis pushes us to Milestone 2, where we choose a specific geographic region to find out the status of restaurants and mobility patterns within specific places in the city and counties. Diagram 7 explores which and how the data is used to complete Milestone 2, 3 and 4.

The project is available in the following repository: [DS4A2020_Empowerment](). We are using .gitignore to list the large datasets and to avoid exceeding the GitHub file size limit of 100MB. The folder data contains subfolders raw (data previous manipulation), interim (files after data cleaning steps and ready to the Exploratory Data Analysis) and processed (data for Machine Learning and Dashboard purposes). Raw is split into 4 categories of data: economics, restaurants, mobility and census. Every dataset is pre-processed as needed and then a clean version is saved in the folder Interim, which has 4 categories as well as Raw. Mobility and Census (raw and interim) are included in .gitignore because the extension of the files. However, the Jupyter notebooks of census data were created considering this issue and we incorporated a data acquisition notebook to download directly the census datasets from the website through web scraping, storage locally and after a first glance of cleaning steps, save the useful structured files in the interim folder. In this way, every user can clone the repository and replicate the process. To get access to the full content of datasets, visit our directory in [Google Drive](). This directory contains exactly the same folders and structure of the GitHub repository, without the storage restrictions.

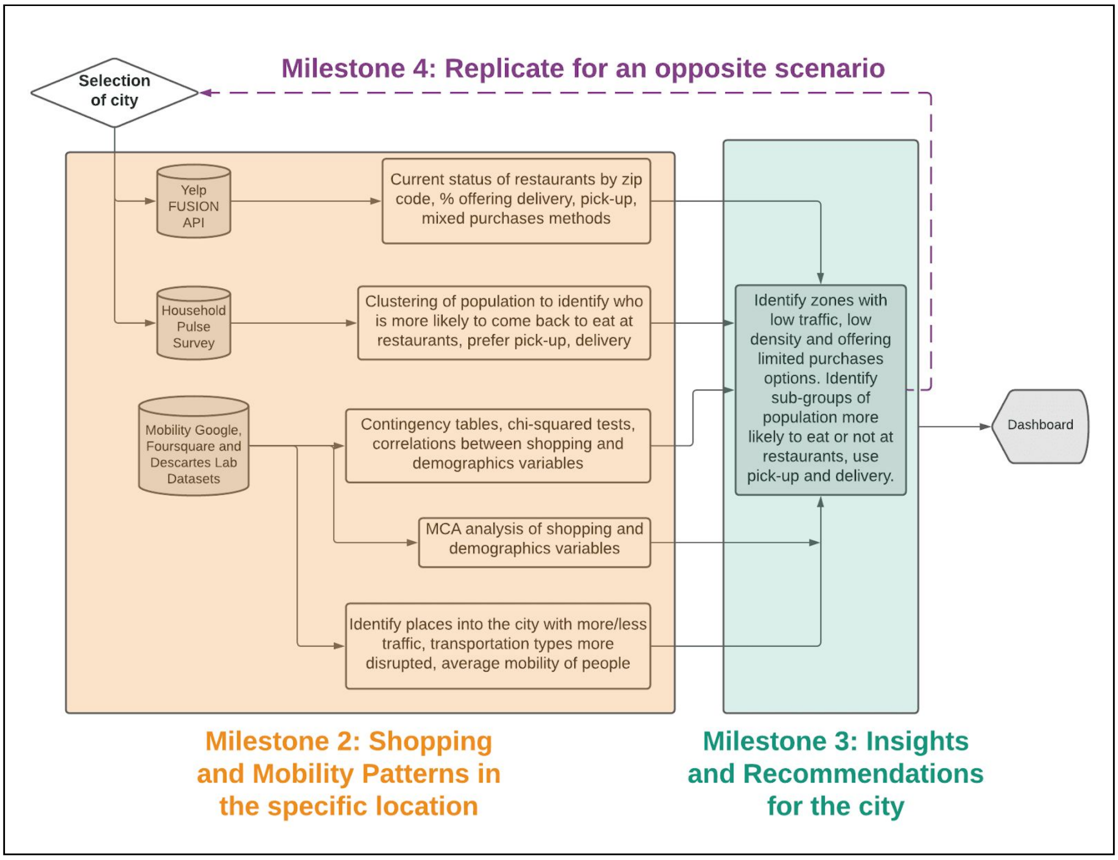


Diagram 7: Structuring Datasets and data processing for Milestone 2, 3 and 4.

Analysis Completed

Part 1 explains the importance of looking at the restaurant industry from a macro to a microeconomic level and includes volatility of GDP by sectors, to understand the key industries driving the contraction and PCE as a catalyst of economic disruptions in the restaurants and employment repercussions in the industry.

Part 2 analyzes the importance of the consumer response to COVID-19 and how that might impact the business, involving dine-in seating reservations during 2020 in different locations, analysis of changes in payment and purchase modality, and willingness to eat at restaurants across the country and by specific demographic variables. Additionally, mobility trends are explored to discover which venues, types of transportations were more affected and tracking of the average mobility of regular members of the communities to evaluate the disruptions in different locations.

Part 3 is an in-depth EDA for two specific locations, using consumer preferences, mobility, and information from the current scenario of restaurants in that specific region to get insights and recommendations about what businesses could do to survive. Besides that, we analyze statistically significant differences between sub-groups of the population and we build a model to segment the population based on their consumer preferences and demographics variables.

EDA Part 1: Economic perspective

1.1 GDP Setting

In order to assess impact, we must consider how GDP has been affected before and after COVID impact (year 2020).

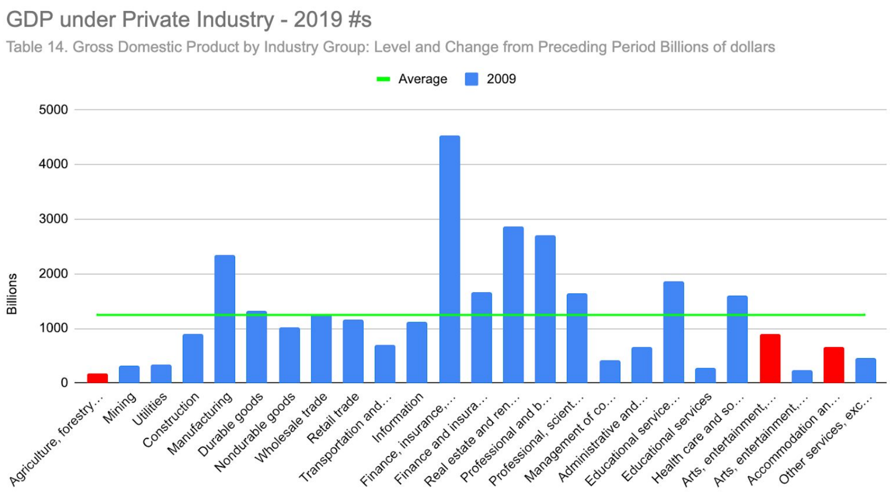


Figure 1.1: GDP under Private Industry 2019

The graph above illustrates the contribution of different industries (under private industry) that impact GDP. We took the liberty to highlight in red those industries that surround restaurants. Industries are not the biggest contributors to GDP (in terms of weight), however, they do encompass a large workforce and business owned operations.

1.1.1 Volatility Over the Years

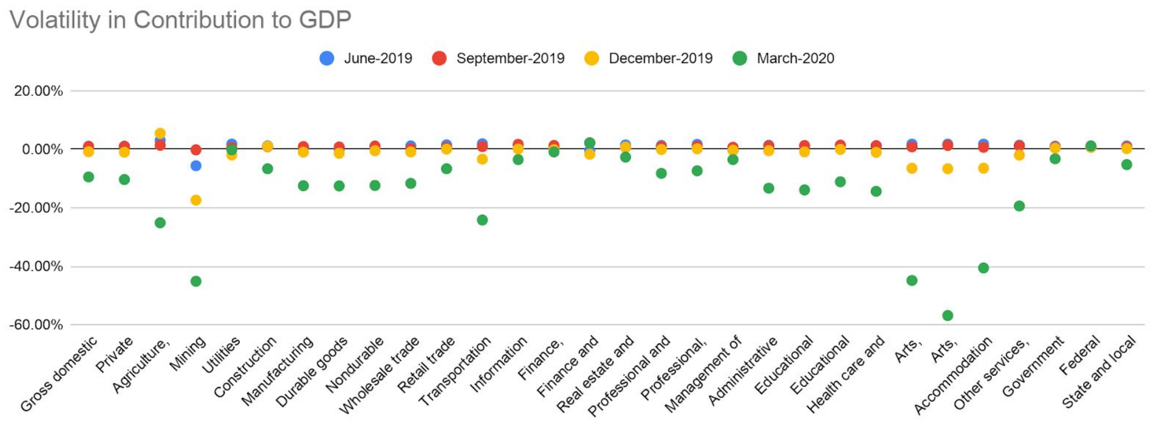


Figure 1.2: Volatility in Contribution to GDP

We noticed a trend in the latter quarters of 2019 and the first quarter of 2020 [see Figure 1.2] where volatility increases significantly. Our hypothesis is that the advent of COVID has greatly impacted GDP in total and clearly certain industries over time.

1.1.2 GDP Impact of Covid

Assessing GDP trends (Figure 1.3), we see that GDP for the US has gone down as a direct result of the largest contributor (in size), Private Industries, going down, even though Government remains flat. We can also notice that the rate at which GDP decreases is directly correlated to that of Private Industry.

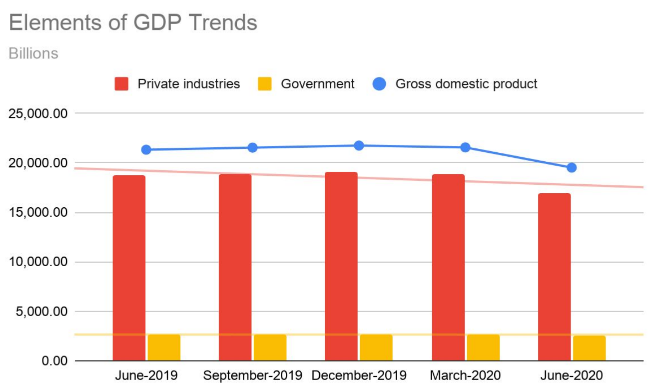


Figure 1.3: Elements of GDP Trends

Looking at the Figure 1.4, we also learn that both Goods and Services were adversely impacted (for their own reasons) as a result of COVID.

Figure 1.4: Goods versus Services

While the percentage changes between Goods and Services is relatively consistent, the capital amounts associated with each, as a contributor to GDP, are vastly different. This is enveloped by the size Services has over the US economy (in terms of $).

1.1.3 Impact of Restaurant Industry

Assessing this, the industries that we are investigating (that in some way are a derivative or main driver of the restaurant industry) are also the ones that have experienced the most volatility, and in particular, losses in latter quarters.

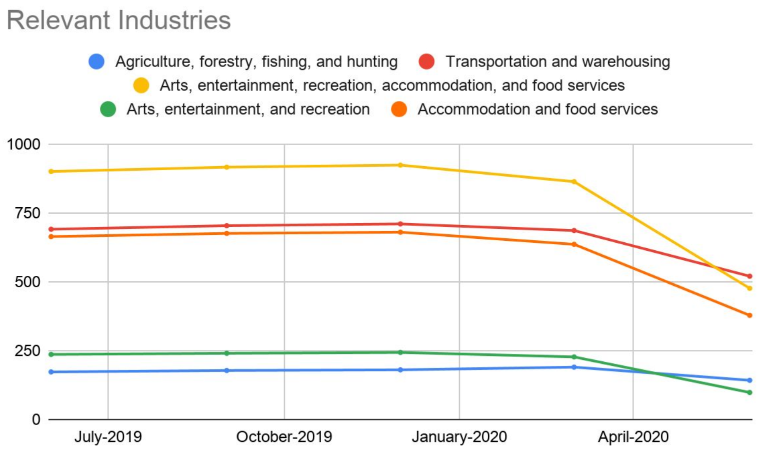


Figure 1.5: Industries related to the Restaurant Industry

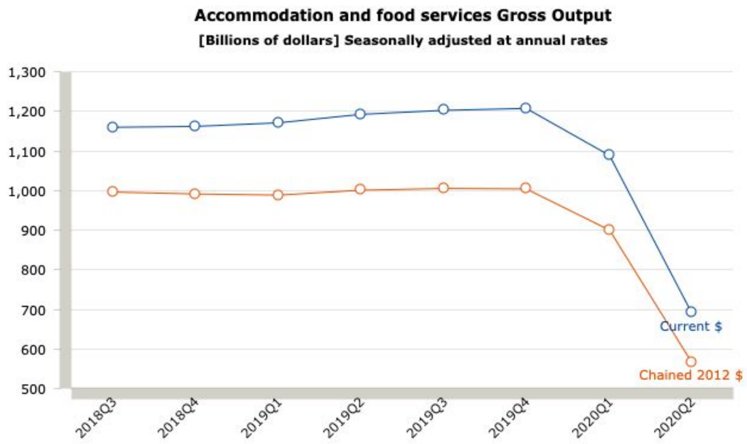


Figure 1.6: Accommodation and food services Gross Output

In 2020Q2, Accommodation and food services nominal gross output decreased to $691.1 billion. Real gross output, adjusted for changes in price, decreased 84.3 percent.

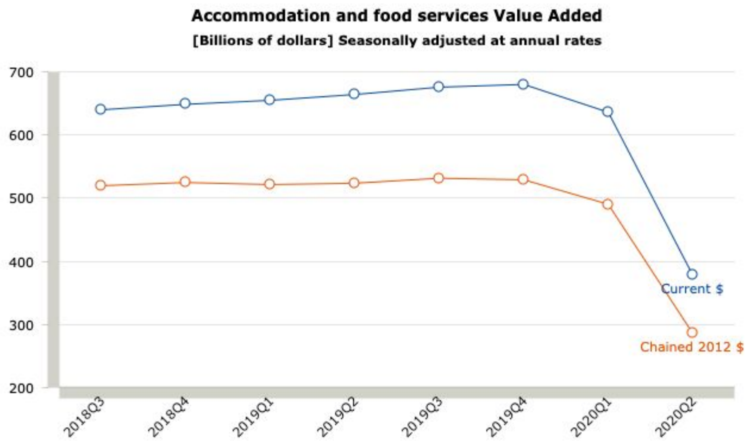


Figure 1.7: Accommodation and food services Value Added

In 2020Q2, Accommodation and food services nominal value added was 1.9 percent of GDP and decreased to $378.1 billion. Real value added decreased 88.4 percent and contributed -4.38 percentage points to the change of -31.4 percent in real GDP.

1.2 Personal Consumption Expenditures

We also wanted to look at changes in Personal Consumption Expenditures (as a whole and by industry). We know that many industries (primarily within the services sector) highly depend on the expenditures of the populace. By studying where this capital (now stressed because of COVID) went. This allows us to assess a fundamental reason why such industries were adversely affected.

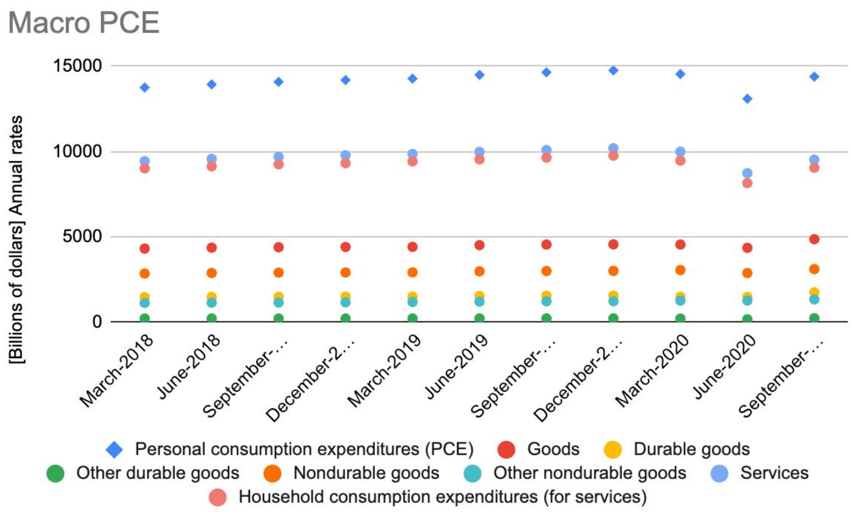


Figure 1.8: Macro Personal Consumption Expenditure 2018-2020

We see that by Q1 of 2020 (March-2020), the total PCE trend had reversed downward hitting the trough at Q2 of 2020 (June-2020) and then recovering to near-normal levels in Q3. The key area that impacted these macro PCE numbers was clearly under Services and Household consumption expenditures (for services). All other areas had a marginal difference and impact on the macro PCE numbers.

We did further analysis on the % growth or decline on a quarter-to-quarter basis for the last three years and then did conditional formatting to highlight key areas of dramatic change (positive or negative). We can clearly see heightened volatility entering 2020 vs any other period over the last three years. See details in Appendix, Section 1, Economic Perspective, Table A1. This prompted us to further explore the areas of greatest movement and impact over the macro PCE numbers. The Figure 1.9 illustrates the degree of volatility for all of the industries driving PCE over time. We learn that not all areas were affected by changing PCE behaviors during COVID. It was chiefly driven by a few (major industries) which ultimately impact the overall PCE numbers and also respective GDP contribution performance.

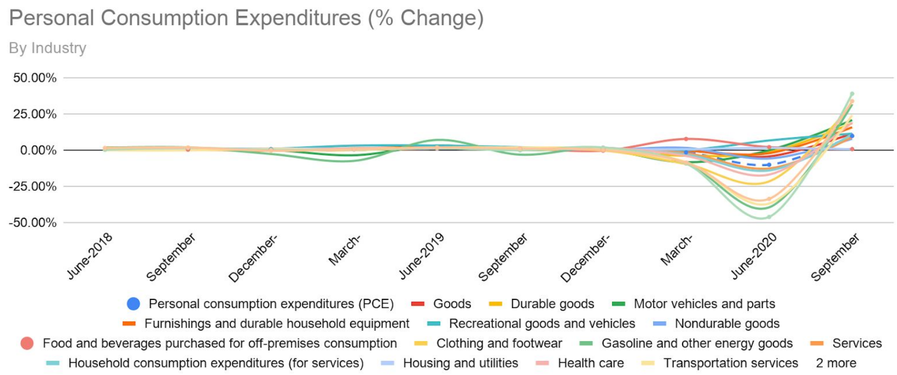


Figure 1.9: Percentual change of Personal Consumption Expenditure 2018-2020

We dig deeper in those industries that experienced the most volatility and note that the volatility didn't all correlate the same way (downward during Q1 and Q2 - 2020). We see that **Final consumption expenditures of nonprofit institutions** and **Food and Beverages purchased for off-premises consumption** actually increased during the early part of 2020 (while the rest fell). This makes sense as there was a lot of fear and a lack of information regarding COVID and best practices to contain it -- we were learning a lot and certain information was being distributed as it came. This prompted the populace to be risk-averse and to contain cash (even amidst a capital injection, "COVID Relief Act"). People chose to purchase and eat more at home, which impacted the restaurant and recreation industries. There was an influx of capital directed to nonprofit institutions; likely helping those in most need or institutions doing research to fight against the pandemic.

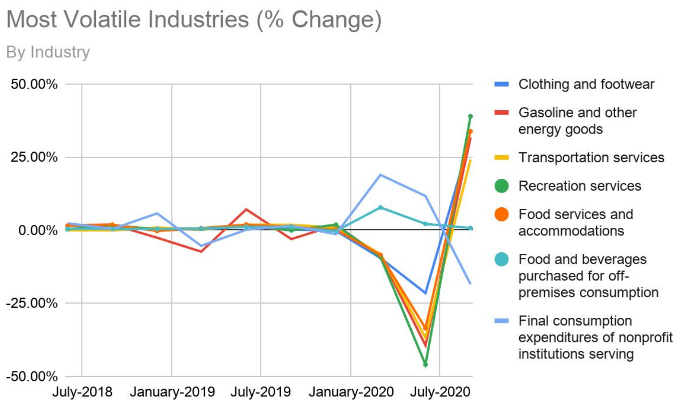


Figure 1.10: Percentual change of PCE by Industry

The following areas experienced the greatest contraction: Recreation Services, Food services and Accommodation, Transportation Services, Transportation Services and Gasoline and other energy goods.

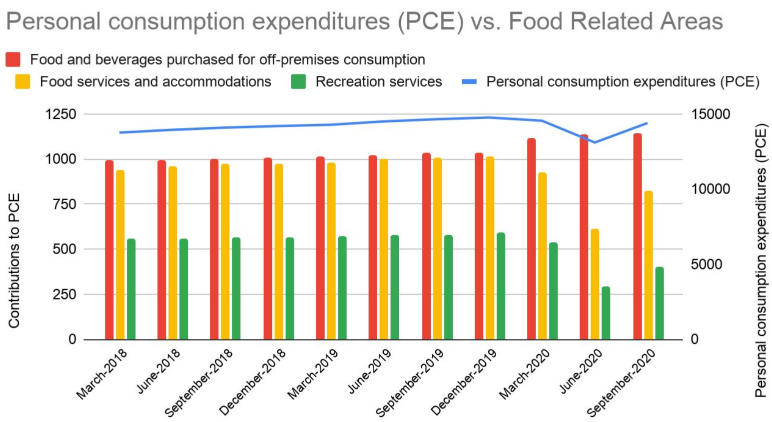


Figure 1.11: PCE versus Food Related Areas

We see in Figure 1.11 that the food-related industries, as a whole, were deeply affected. When you compare food-related industries vs. other bundles (e.g. energy, clothing, finance, etc.), it presents the greatest divergence as a result of consumer expenditure behavior. To illustrate this divergent relationship, we run regressions between **“Food and beverages purchased for off-premises consumption” (i.e. groceries for the home)** vs. **“Recreational Services” & “Food Services and Accommodations” (i.e. restaurants/bars)**.

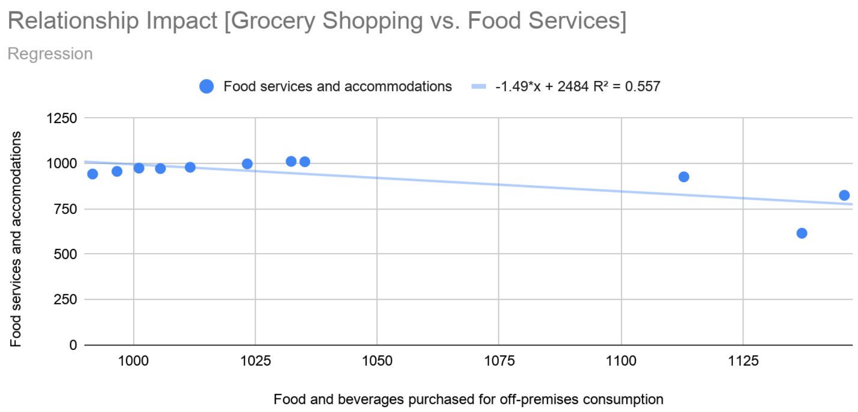


Figure 1.12: Regressions between recreational services and food/beverages consumption

By looking at the regressions above, we see that there’s a slight inverse relationship between both. As more groceries are made, recreation and food services suffer. We can also see that when factoring percentage changes (per quarter) over the last three years between **Food Services and Accommodations** vs. **Food and beverages purchased for off-premises consumption (groceries)**, we notice volatility heightened (both positive and negative) over the last three quarters only. They are illustrated by the clear outliers (-33% and +33% for **Food Services and Accommodations**). whereas the usual relationship is relatively consistent and stable (cluster).

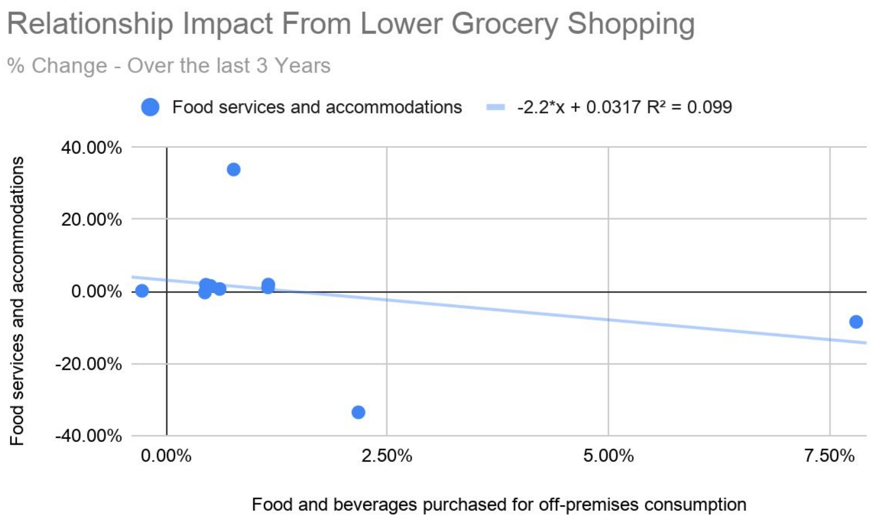


Figure 1.13: Regressions between recreational services and food/beverages consumption over the last three years

1.2.1 Food industry inventories and sales (1992-Present)

Measurements of sales combined with insight on inventory levels speak to the supply and demand of industries.

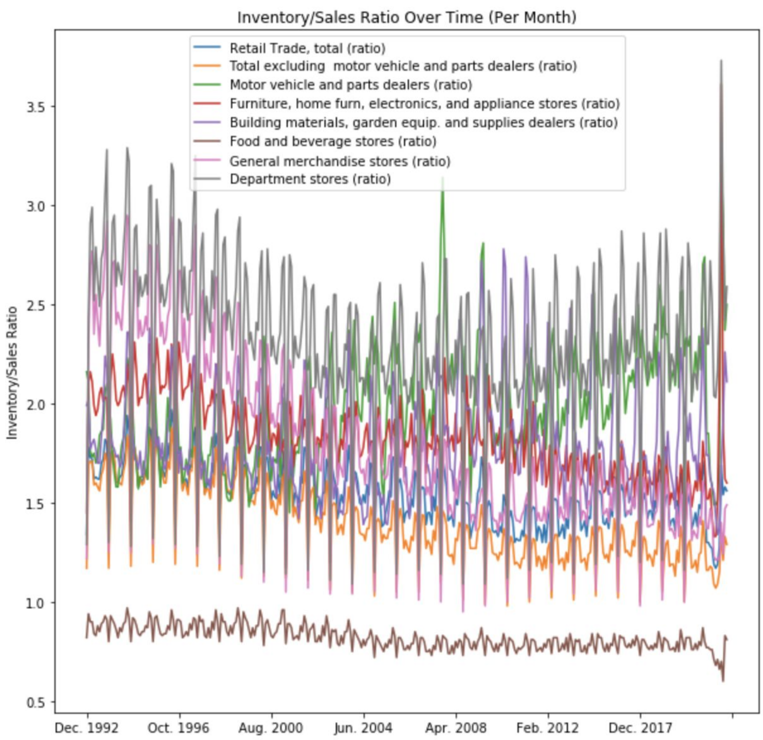


Figure 1.8: Inventory/Sales Ratio of Major Industries (Millions USD; 1992-Present)

Increases in ratio of inventory to sale means less demand for a product. Conversely a lower ratio means scarcity which can mean a hot industry, but also a dying industry. We see from the graph below that the seasonal variabilities are heavily overshadowed by the latest months. The brown line at the bottom represents the sudden decline of the food industry.

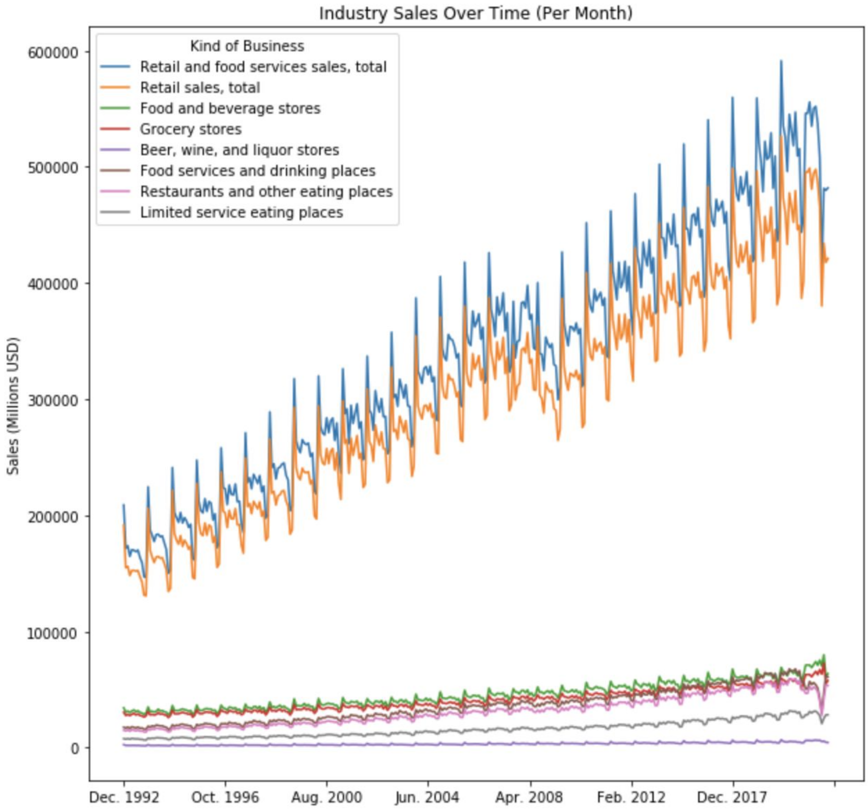


Figure 1.9:Total Sales of Food Industries (Millions USD; 1992-Present)

The above plot shows the total sales of food industries from 1992 to today. The pink line in the bottom half represents the restaurants and other eating places. While all industries took a hit in the early 2020, this one took the hardest, returning to 1992 levels. This collapse had widespread effects as the restaurant industry is a major source of employment.

1.3 Employment Impact

1.3.1 Unemployment Data

The coronavirus continues to deliver unprecedented economic numbers and so we continue our analysis by looking at unemployment. It is no secret that some industries fared much better than others in the pandemic. In the upcoming section, we seek to answer a few questions. How has the restaurant industry fared compared to other industries? How do the numbers change and fluctuate? How do the numbers compare to the Great Recession?

For the upcoming graphs, we used unemployment data from the St. Louis Federal Reserve Bank which tracked national monthly unemployment numbers by industry from 2005 to 2020. We included this range to compare the present coronavirus recession to the Great Recession following the 2008 financial crisis, the most recent recession. The first thing to address is that this Dataset did not have restaurants as its own category. Restaurants are lumped into “Leisure and Hospitality”(L&H) which also includes other service-based sectors such as bars, theme parks, and hotels. All of these sectors have been hit very hard by the coronavirus. Another is that unemployment seen for a specific month is taken on the 1st of that month rather than an average of the unemployment in that month. For example, April unemployment is measured on April 1st rather than being an average of unemployment for April. So unemployment for a specific month may be more reflective on how the previous month ended rather than a general view of the entire month.

The first graph shows the monthly national unemployment in all sectors compared to L&H. The first thing to note is that L&H is more volatile and has on average higher unemployment, even during times of prosperity. From the graph, we can see the seasonal change in unemployment with winter months having higher L&H unemployment compared to summer. L&H was severely hit during the pandemic with unemployment peaking in April at 39.3% while national unemployment stood at 14.4%. April is when many Americans got laid off or furloughed and thus the numbers show this. This is especially seen in the L&H unemployment where there is nearly a 25% difference. This is bigger than the difference in peak unemployment in January 2010 at 14.2% (L&H) vs 10.6% (Total) during the Great Recession. After this spike, the numbers quickly level off for both categories. The beginning of October shows 16.3% unemployment in L&H versus 6.6% unemployment for all categories. This is a remarkable improvement from April but shows just how far L&H has to go. We don't expect the numbers to get that much better as we enter the third wave.

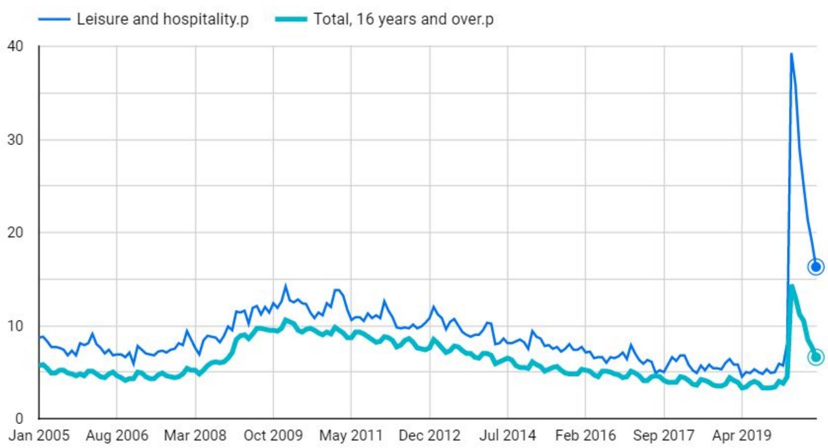


Figure 1.14: Monthly unemployment in all sectors compared to L&H.

The second graph below shows monthly national unemployment in the sectors of Leisure and Hospitality (L&H), Professional and Business Services (PBS), Durable goods, and Non-durable goods from 2005 to 2020. We can see from the graph that L&H still eclipses the other sectors by an absurd amount. With L&H at 39.3% unemployment in April, the next highest sector was Durable goods with 15.1% unemployment. In January 2010, the last period of mass unemployment, there was very little difference between the unemployment of L&H (14.2%) and Durable goods (14.1%). Part of this can be attributed to the fact that the Great Recession was much more of a traditional recession brought on by a financial crisis rather than a pandemic. PBS and Non-Durable goods had a similar peak in April unemployment at 9.8% and 10.2% respectively. In the following months, all industries shown begin to recover fast as more restrictions across the country are lifted. Durable goods by October actually has the lowest unemployment rate (4.6%) of all the sectors chosen. Again we expect the numbers to change once unemployment data for November and December begins to come in.

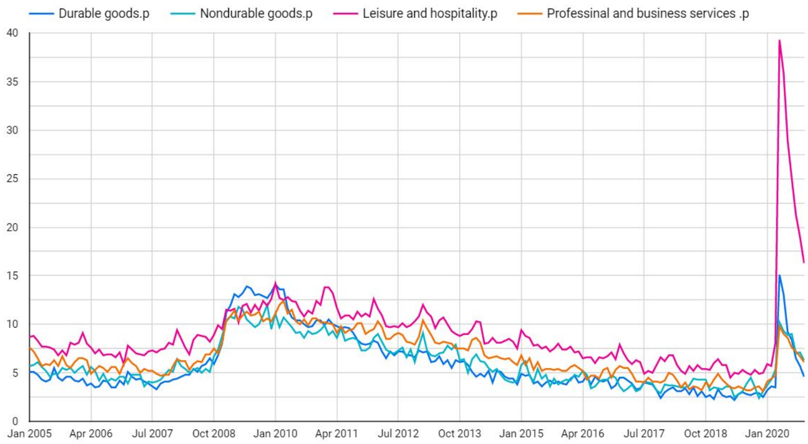


Figure 1.15: Monthly national unemployment in the sectors of Leisure and Hospitality (L&H), Professional and Business Services (PBS), Durable goods, and Non-durable goods from 2005 to 2020

The historical records of employment in Food Services and Drinking Places Industries can be found in Appendix, Section 1, Economical Perspective, Figure A1.

EDA Part 2: A consumer perspective

2.1 Consumer Behavior: Dining-in impact

2.1.1 Seating Dining Data

The pandemic has had an enormous amount of impact on restaurants but amongst all the mayhem it can be hard to tell how much effect this has had on daily operations. In this section, we review Seating Data from Restaurants around the world, including the United States. This dataset tells us the percentage change in how many seats are occupied from the baseline of February 18th, a few weeks before the pandemic, all the way to November 24th. This can help us answer the questions of how many daily operations were impacted? How have restaurants in certain cities or states fare against each other? Are there any discernible patterns we can see?

The figure shown below shows the relative change in seating data from the baseline value (February 18th) to today. We are comparing the states of California, Texas, as well as the US, in general, to track how seating data has affected these states individually. All three closely mirror each other up until May 1st where restaurants were beginning to open in Texas as well as the rest of the US. California restaurants started to resume operations during memorial day weekend (May 23rd). California and the rest of the United States closely mirror each other for the rest of the graph with California lagging slightly behind. Texas had a much more aggressive reopening strategy and the graph shows this with Texas consistently having a higher percentage of people seated compared to the US and California. Texas even manages to return to baseline seating on Labor Day weekend (September 6th). However, these gains are short lived. As the trend line shows, seating begins to decline after Halloween weekend (October 29th to 31st) as the US third wave begins to pick up speed.

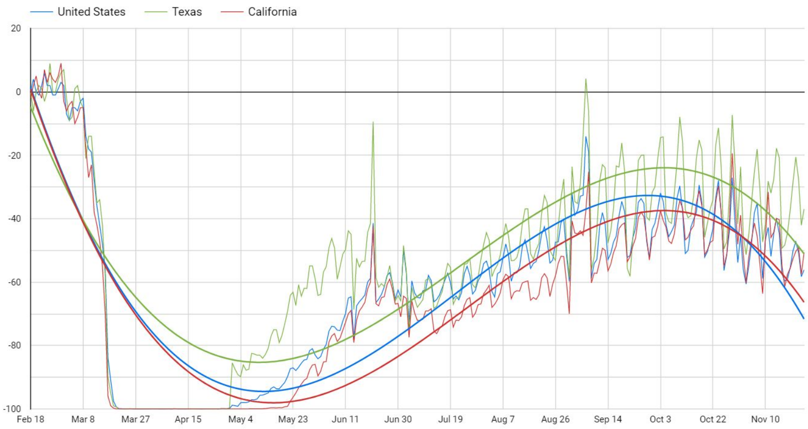


Figure 2.1: Relative change in seating data from baseline value in Texas, California, and Nationwide

Next we look closer at the state of California which is famously or infamously known for its strict coronavirus regulations. We compare the state of California, the cities of Los Angeles and San Francisco, as well as the US as a whole. This can give us clues as to how urban areas and their restaurants fared since Urban areas tended to have even more strict coronavirus regulations. The graph below shows that large California cities and their restaurants tended to be worse off compared to the rest of the country. Both LA and SF opened up more cautiously and experienced their first modest increase in seating on June 21st. For the rest of the pandemic, these cities continue to have modest increases in seating but still never being able to achieve anything close to baseline seating. Like everybody else though, the cities experience a great decline in seating following Halloween weekend (October 29th to 31st) and into the third wave.

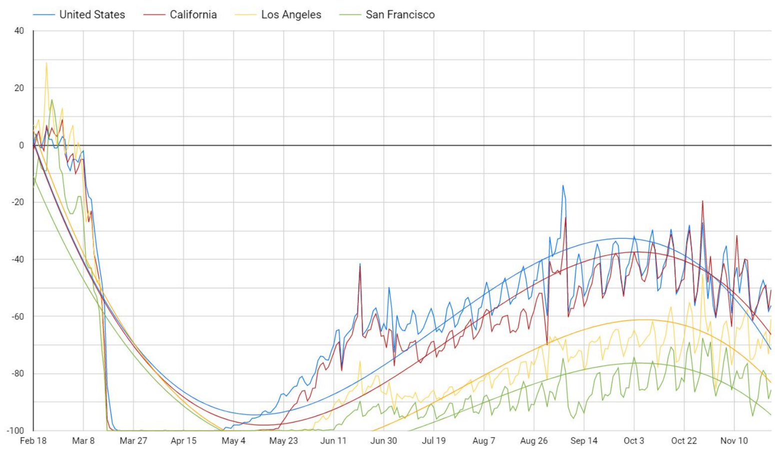


Figure 2.2: Relative change in seating data from baseline value in California, Los Angeles, San Francisco, and Nationwide

2.2 Overall changes in consumer behavior

In this section, we focus on Phase 2 and 3 of the Household Survey, which started on the 13th week of the study onwards and contains variables related to changes in shopping behaviors. Firstly, we explore some background information on the composition of the sample population during the weeks covered in this analysis (weeks 13th-20th of the survey, from August 19th to December 7th, 2020):

- **Region**: 32.8% of the surveyed people belong to the West Region, followed by 31.1% who live in the South; 20.5% in the Midwest, and 15.5% in the Northeast.
- **Gender**: **59% of the surveyed people are identified as women** (binary options only).
- **Race: 82% identify themselves as white**; 7.8% as black; 5% as Asian and 5% other race or race in combination.
- **Educational attainment**: **29.2% of the surveyed people have a Bachelor's degree across the country,** followed by 24.8% who have completed a Graduate's degree and 21.6% an incomplete college degree. Only 2% have less than a high school or an incomplete high school.
- **Generation**: 30% of the surveyed people belong to Generation X, 25% to Millennials, 20% to Generation Jones, 15% to Baby Boomers, 6.7% to Silent Generation, and 3% to Generation Z. It means that the **majority of the sample was born between 1946 and 1996**.

Population by Region, Racial, and Educational attainment variables can be found in Appendix, Section 2, Consumer Perspective, Figure A.2. From there, we conclude that the distributions of race and educational attainment are similar across the regions, except by two observations: in the South, black people represent 15%, more than double that in the rest of the regions and in the West, other races and Asian people are the 15% in combination, more than the double that in the other places. Besides that, 75% of the survey respondents don't have children and 70% reveal to have at most 4 members in the household. Details in Appendix, Section 2, Consumer Perspective, Figure A3.

Changes in shopping behaviors

The survey asks respondents about changes they made in the last 7 days in their shopping behaviors according to purchase/payment modalities and eating at restaurants. Figure 2.3 reveals the percentage of people who declared did **more purchases using one of the specified modalities during the last 7 days**: made more purchases online, more purchases by curbside pick-up, both or more purchases in-store.

In the graph below, we see the cumulative percentage of change over time to explore the trends over the three modalities. **Nationwide, online purchases are highly preferred** and they increased across the weeks as an alternative to in-store and curbside-only purchases, which were decreasing over time. At a granular level, the preferences by educational attainment, generation, and race are exposed in Figure 2.4. **By educational attainment, online purchases are widely preferred over in-store and pick-up purchases**, especially in higher educational levels. By generation, we see that more than 10% of boomers, silent and jones respondents favored in-store purchases and by race, more than **20% of black people and other races increased their in-store and pick-up purchases**. By region, **17% of respondents in the South and Midwest revealed made more in-store and pick-up purchases**. In the **West and Northeast, at most 13% made the same choices**.

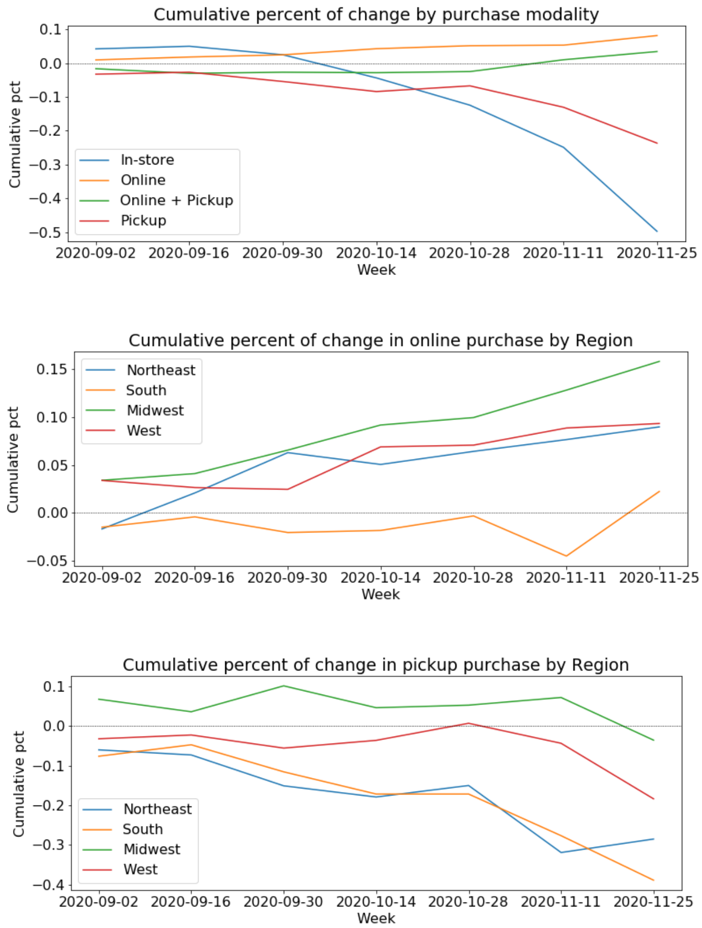


Figure 2.3: Cumulative percentage of changes in purchase modality

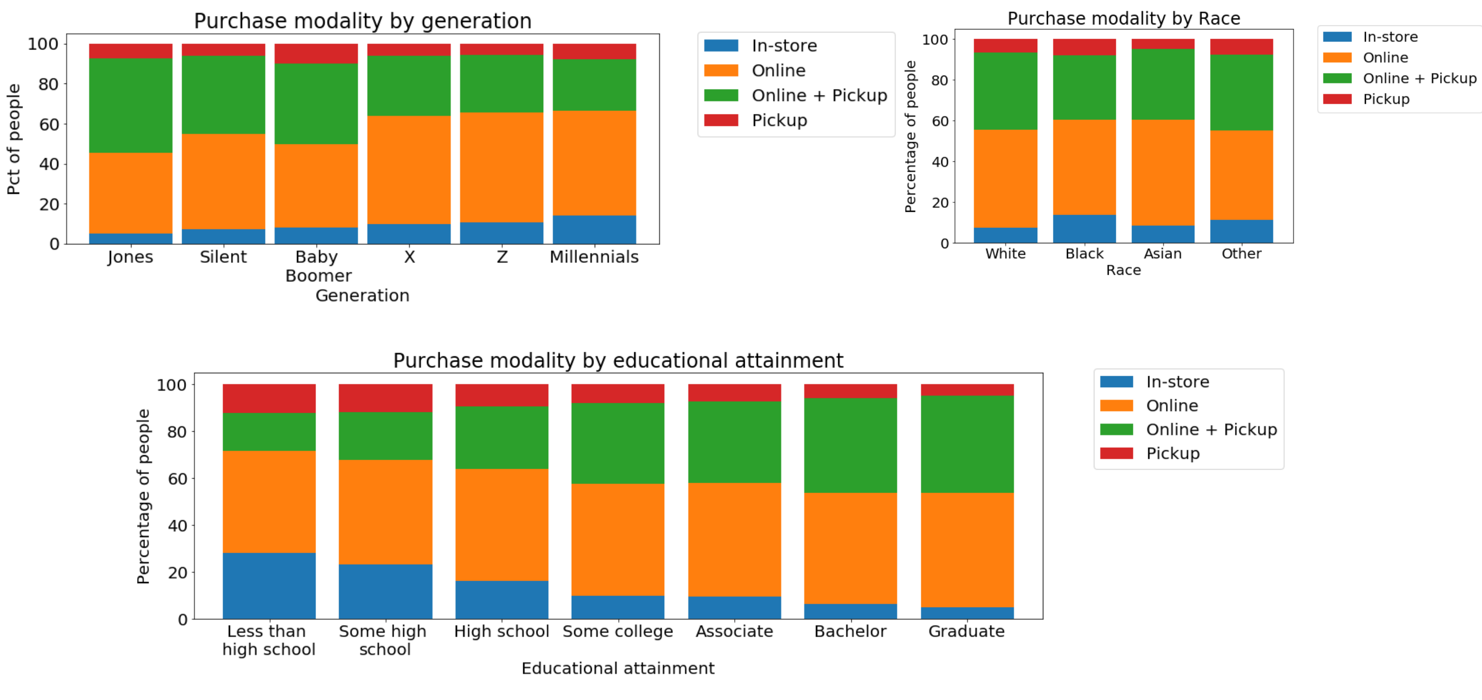


Figure 2.4: Purchase modality by educational attainment, generation, and race

The same analysis was made for payment modality. The surveyed participants were asked about **their payment changes during the last 7 days**. By region, **8.2%** of surveyed participants from **the Midwest** used **more cash last week** and **6.9%** made the same payment choice **in the Northeast. West and South, was 7.3 and 7.5%** respectively.

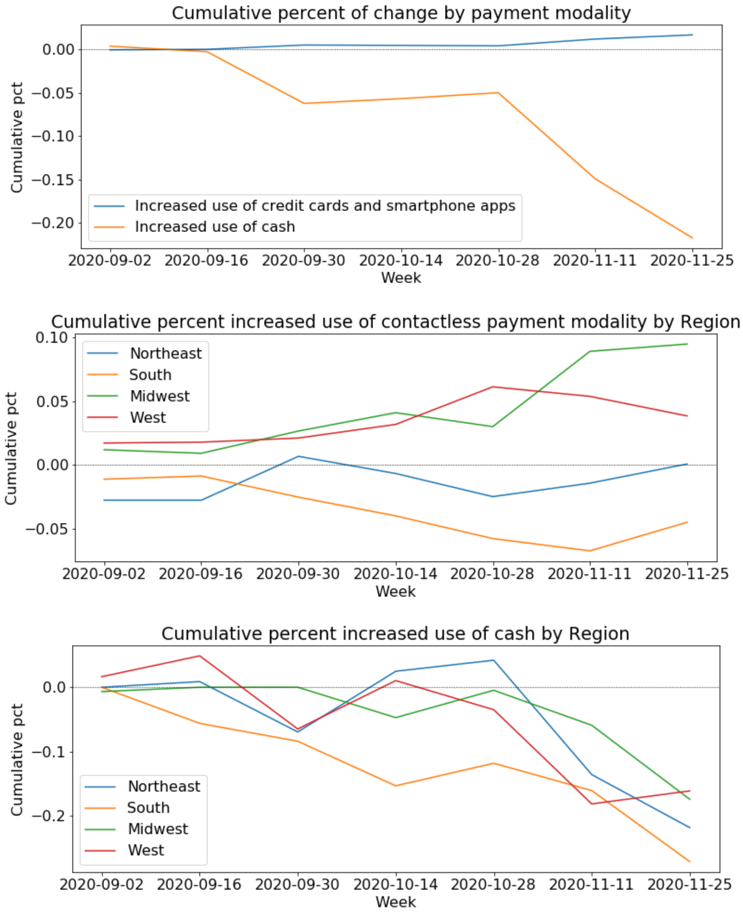


Figure 2.5: Cumulative percentage of changes in payment modality

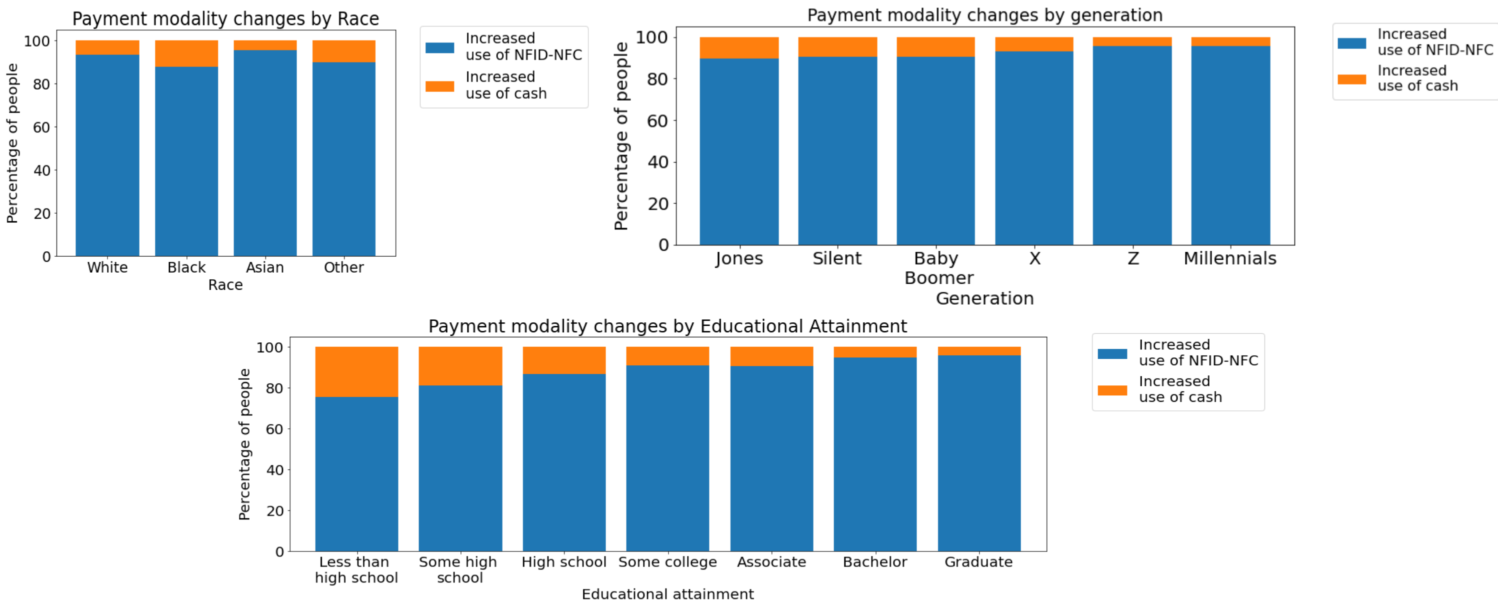


Figure 2.6: Purchase modality by educational attainment, generation, and race

And finally, the survey asked if people **increased their resumes or avoided eating at restaurants during the last 7 days**. According to Figure 2.7, the percentage of people who resumed eating at restaurants was decreasing over the weeks. Across the country, more than **86%** of the sample population **avoided eating at restaurants**, with the **highest % in the West region (88.4%).** The West experienced the highest resume for eating at restaurants during Fall 2020 when the outdoor dining and indoor at 25% of capacity was implemented, but in November, the indoor modality was forbidden again and more people avoided eating at restaurants.

As exposed in Figure 8, **elderly much more avoided eating at restaurants than young people,** presumably because of the vulnerability and risks that the pandemic means for older people. By race, **black avoided more eating at restaurants, followed by Asians and race in combination**.

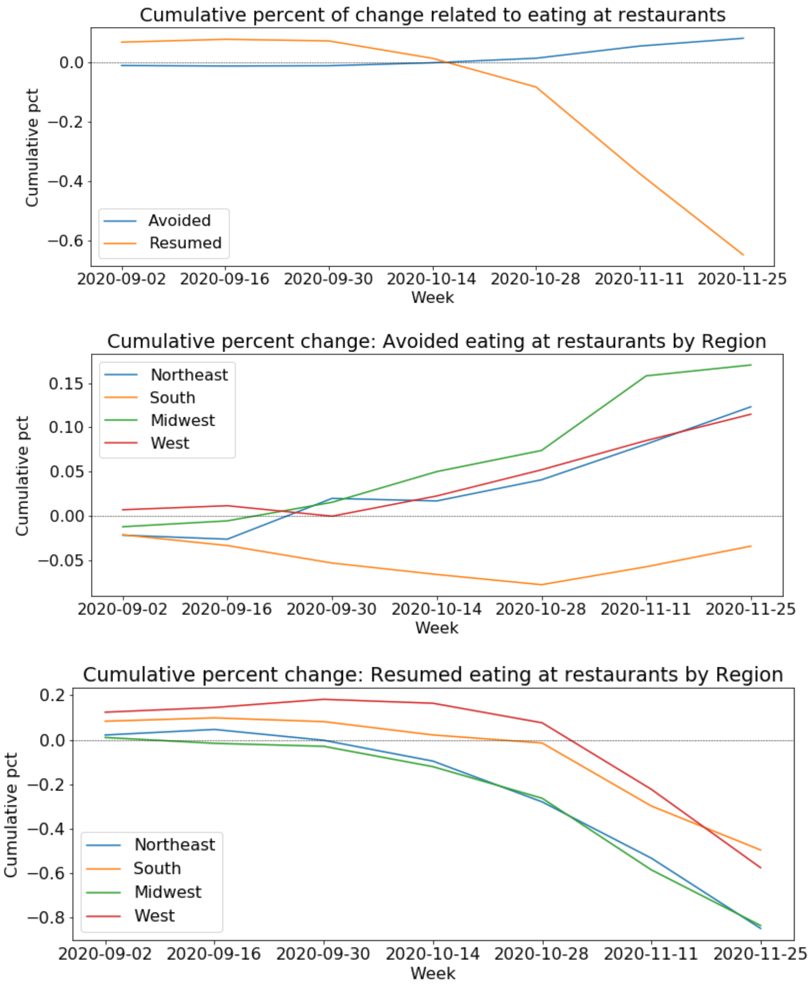


Figure 2.7: Cumulative percentage of changes in eating at restaurants.

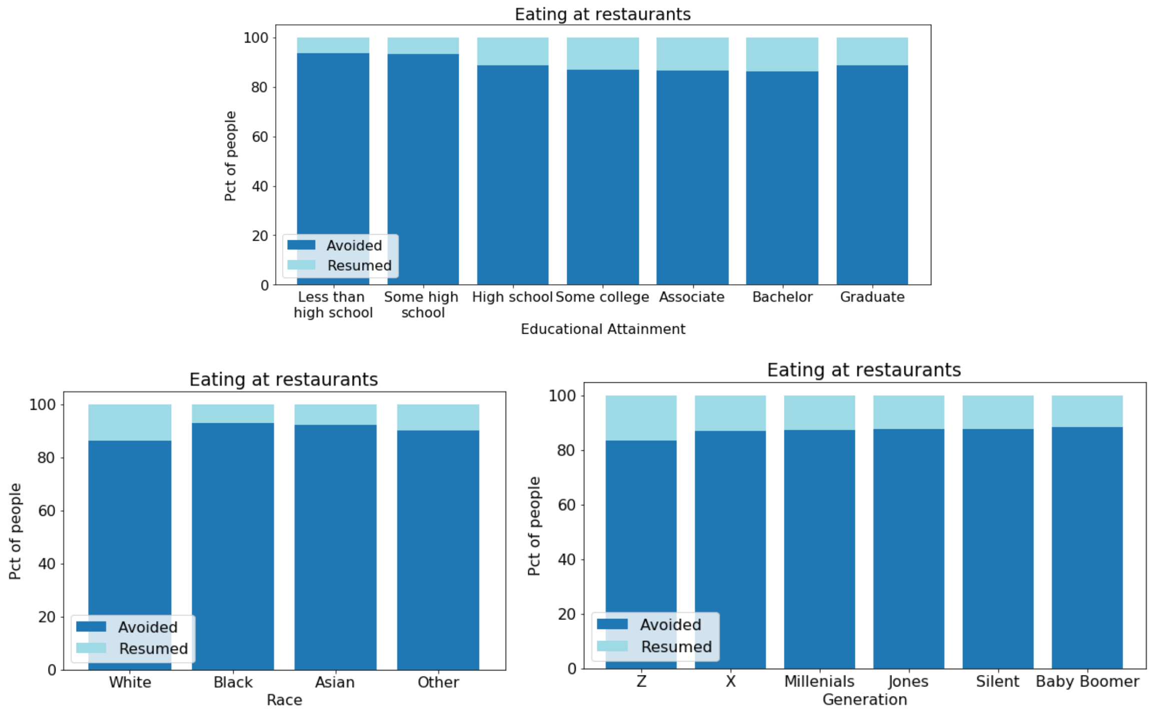


Figure 2.8: Resumed eating at restaurants by educational attainment, race and generation.

2.3 Mobility Changes in Response to COVID-19

2.3.1 Apple Mobility Reports

How much has changed the number of route requests that people have made during this year? How the transportation types have been affected during the year with regard to the baseline volume? Based on the reference volume of requests, can we identify states more/less disrupted in terms of weekly and monthly routes requests?

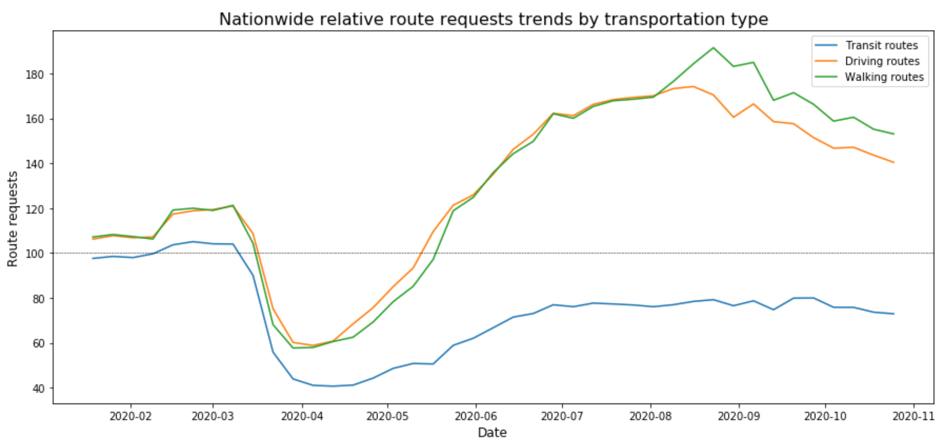


Figure 2.9: Route requests trends by type of transportation

Figure 2.9 explores the nationwide directions requests over the year by type of transportation relative to a baseline volume (January 13th, 2020), consistent with normal. The overall route requests dropped significantly with the lockdowns across the country, but from April, driving and walking requests increased at the same strong rate, especially during the summer season. However, both declined from the spring. Transit experienced a slow increase during the autumn to keep practically constant from the summer onwards.

Details about the distribution of requests by states during 2020 can be found in Appendix, Section 2, Consumer Perspective, Figure A4.

Then, by state, we can identify the states more disrupted (who boxplots are located so far off the baseline, which is a reference point of normal requests by type of transportation and it's related to every particular state). Based on Appendix, Section 2, Consumer Perspective, Figure A5, Hawaii, Washington, New York, California, Oregon, Illinois, and Massachusetts, to name a few, are the most disrupted states by changes in-transit transportation. Some of them have high variability over the year, like New York and Massachusetts, compared to California, Hawaii, or Washington. On the right, only Kansas, Oklahoma, Idaho, New Hampshire, South Carolina, Arkansas, Alabama, and Mississippi locate their boxplots (and precisely the mean of the transit requests) over the transit baseline.

Driving and walking requests are over the baseline in almost all the states, except in Hawaii (Appendix, Section 2, Consumer Perspective, Figures A6, A7). Maine, Idaho, South Dakota, and Wyoming are furthest from the baseline in both types of transportation. Florida, New York, California, Massachusetts, and Louisiana keep closest (and even) under the baseline in driving and walking directions requests.

2.3.2 Foursquare Foot Traffic

How much has changed the foot traffic in food venues (grocery, convenience, discount stores versus fast food and casual dining restaurants)?

In Figure 2.10 we explore the foot traffic in venues where people can get food: stores (convenience, discount, big box, grocery stores) and restaurants (casual dining and fast-food restaurants). The plot below shows how all the venues, except for casual dining, have kept around the baseline foot traffic (even higher between March and April).

Then, in Figure 2.11, the same analysis exposes the abrupt drop in foot traffic in the casual dining venue in April across all the regions of the country. From May the venue is recovering in all the regions, but they are still far away from the baseline. Additionally, **foot traffic has been slightly higher in the South and Midwest than in the Northeast and West**. On the other hand, the **Fast Food venue was also disrupted but on a smaller scale than casual dining**. Finally, in Figure 2.12 we go deeper into the West region and compare the foot traffic at the Fast Food venue in the metropolitan areas Los Angeles, Seattle-Tacoma, and SF Bay Area. New York, from the Northeast, is also included. **Los Angeles and the SF Bay Area are the most affected areas so far.**

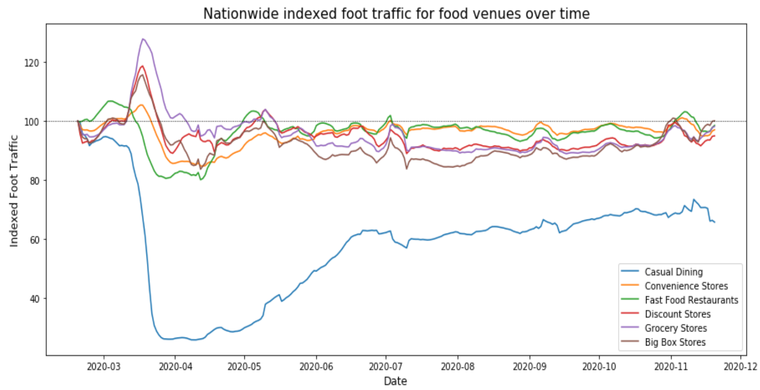


Figure 2.10: Nationwide indexed foot traffic in food venues during the year

Figure 2.11: Indexed foot traffic in Casual Dining versus Fast Food Restaurants during the year by Region

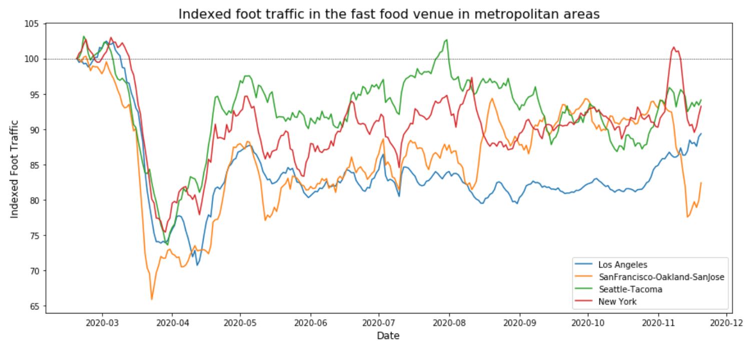


Figure 2.12: Indexed foot traffic at Fast Food Venues in 3 metropolitan areas from the West (Los Angeles, Seattle-Tacoma, and SF Bay Area) and 1 metropolitan area from the Northeast (New York).

2.3.3 Descartes Lab Data

From the previous sections, we collected information about mobility trends related to route requests and foot traffic into venues, but, how much has changed the distance that people move daily? This data shows how the max distance mobility has changed compared to a baseline, defined as the median of max distance mobility measured during 2020-02-17 to 2020-03-07. Data provides a daily and by state distance that a typical member of the given population moved (in km) and indexed distance over the daily baseline. From Appendix, Section 2, Consumer Perspective, Figures A8, A9, can be seen that the **median mobility in states from the Northeast and West are lower than the distance mobility in the South and Midwest regions**. According to Figure A8, **Wyoming, North, and South**

Dakota are the only states whose median indexed mobility is over the baseline. They also have the **highest level of variability of mobility over the year.** Instead, **California** is **the furthest state from the baseline and the daily distance in this sample population doesn't exceed 4 km** (Appendix, Section 2, Consumer Perspective, Figure A9).

2.3.4 Summary Consumer behaviors and Mobility Data

- West and Northeast regions are recovering slowly from their regular foot traffic, related to the South and Midwest.
- The sample population in the Household Survey that prefers in-store and pick-up purchases is higher in the South and Midwest than in the West and Northeast.
- States from West and Northeast also have the most disruptions on transit, driving, and walking directions requests.
- Casual dining is the most disrupted food venue during the year across the country.
- Split by region, the food traffic recovering in the casual dining venue is slightly slow in the West and Northeast than in the South and Midwest. The West is also the region with slow foot traffic recovering in the fast-food restaurant venue.
- West and Northeast are the places with higher use of contactless payment methods.
- West is the region where more people resumed eating at restaurants during week 13th and 20th from the Household Survey.
- Comparing LA, SF Bay Area, Seattle and NY, Los Angeles and SF Bay Area are the most affected areas so far in the indexed foot traffic at the fast-food venue.
- Median mobility in states from the Northeast and West are lower than the distance mobility in the South and Midwest regions.
- California is the furthest state from the mobility baseline and the average daily distance doesn't exceed 3 km.

EDA Part 3 (in-depth): A case study for specific locations

The following analysis explores the **customer preferences** and the **restaurant situation in a specific geographical location**.

Section 1 displays lower triangular correlation heatmap matrices of shopping behavioral and demographic variables, as well as the relationship between the different shopping variables, consumer profiles using clustering techniques and visualization of variables in 2D through Multi Correspondent Analysis (MCA), an alternative to PCA for categorical data.

- **Pearson correlation** was used to compare the shopping variables because of their binary nature. Since the demographic variables are categorical, we proceed to use two alternative methodologies:
    - **One-hot encoding** the demographic variables to calculate the Pearson correlation. That was applied to get the lower triangular correlation heatmaps arrays.

    - **Cramer's V**, a measure of association (**chi-squared statistic**) between two categorical variables.
- Consumer profile: **Clustering** of population based on demographic and shopping behaviors using **k-modes**. **Silhouette score** is performed to determine the best number of clusters.
- **MCA dimensionality reduction** to visualize demographics against shopping variables and data points (sample population) split by the clusters from consumer profiles.

Section 2 examines the mobility trends in the selected location. Some of the metrics included are the average distance that a typical member of the community moves on a regular day, mobility around different venues, and route requests to drive, walk and transit into the city. In this section we apply additive decomposition of time series in some specific examples, to remove seasonality and noise from the signal and preserve only the trend, which is the significant signal component for the current purpose.

Section 3 investigates the restaurant scenario in the city: rankings, prices, popularity, locations, and operations details (the type of transactions allowed).

Finally, using consumer preferences, mobility trends, and restaurant features, we extract recommendations by zip code.

3.1 San Francisco Bay Area

Since California is one of the states with more disruptions in mobility patterns (maximum distance for a typical member in a regular day, foot traffic in casual dining, fast food restaurants, and drastically hit in use of public transportation and driving route requests), this section is going to focus in one particular metropolitan area, San Francisco Bay Area, the location **doing lower in-store purchases** (5.8%), the **second-largest online purchase** (58.5%), using the **highest contactless payment methods** (95.2%) and **leading the avoidance** (voluntarily or not) **of eating at restaurants** (92.7%).

3.1.1 Consumer preferences

Defining **protective** behavioral change as any conforms to pandemic-avoidance behaviors (e.g., increasing online shopping, avoiding eating at restaurants), whereas a **relaxing** behavioral change as any weaker adherence to pandemic-avoidance behaviors (e.g., increasing in-store shopping, resuming eating at restaurants), it's clear than protective behavioral changes are strongly correlated with each other and negatively correlated with relaxing behavioral changes. **Online purchases** are related to **pick-up**, **contactless payment methods,** and **avoided eating at restaurants**. **In-store purchases** are related to **increment in the use of cash** and **resumed eating at restaurants** (Figure 3.1, left).

The variables more correlated with avoidance of eating at restaurants are educational attainment, marital status, and generation (Figure 3.1, right). To dive a little further into these results, the correlation between **increased avoidance of eating at restaurants** and the demographic categorical variables in their one-hot encoding version is calculated, as is exposed in Appendix, Section 3.1, San Francisco Bay Area, Figure A10. The lower triangular correlation matrix of **avoidance of eating at restaurants** shows a slight relationship between the analyzed shopping variable and marital status 1 (married people), high educational attainment (Master and Ph.D.), and Generation X and Jones. The same analysis for **increased resume eating at restaurants** can be found in the Appendix, Section 3.1, San Francisco Bay Area, Figure A11, as well. In this case, there is a relationship resume eating at restaurants, race (white), high educational attainment (college degree) and Millennials. Additionally, the lower triangular matrices show the relationship between demographics for the population studied (race-generation, race-education and education-generation).

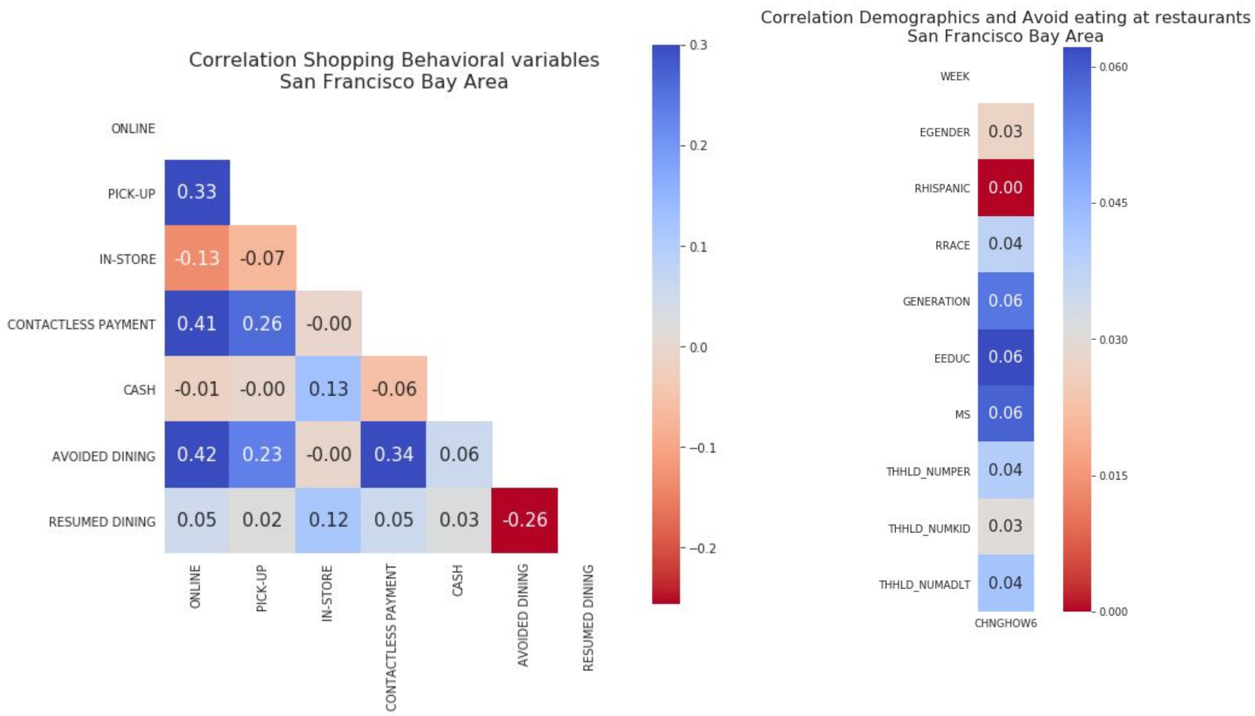


Figure 3.1: Lower triangular correlation matrix of shopping behavioral variables (left) and categorical association of demographic variables with avoidance of eating at restaurants (right).

The conclusion for the same analysis in the rest of the shopping variables are listed as follows:

- White people are more likely to resume eating at restaurants.
- The use of cash and in-store purchases is more related to the black and Hispanic communities and race in combination, as well as educational attainments from 1 to 4 (no college or more advanced degree) and people from the Baby Boomers generation, generation Jones and Silent generation.
- The increase in online and pick-up purchases is more related to Millenials, Generation X, and higher educational attainments (Master and Ph.D.).

To complete this first in-depth subsection, we build consumer profiles using the consumer preferences and demographics variables through **K-modes. Silhouette analysis for K-modes** is used as reference to find how many clusters are necessary to separate classes the maximum possible. K=4 is selected. The distribution of demographics and shopping variables by cluster can be found in Appendix, Section 3.1, San Francisco Bay Area, Figures A12, A13. Essentially, the main features by cluster are the following:

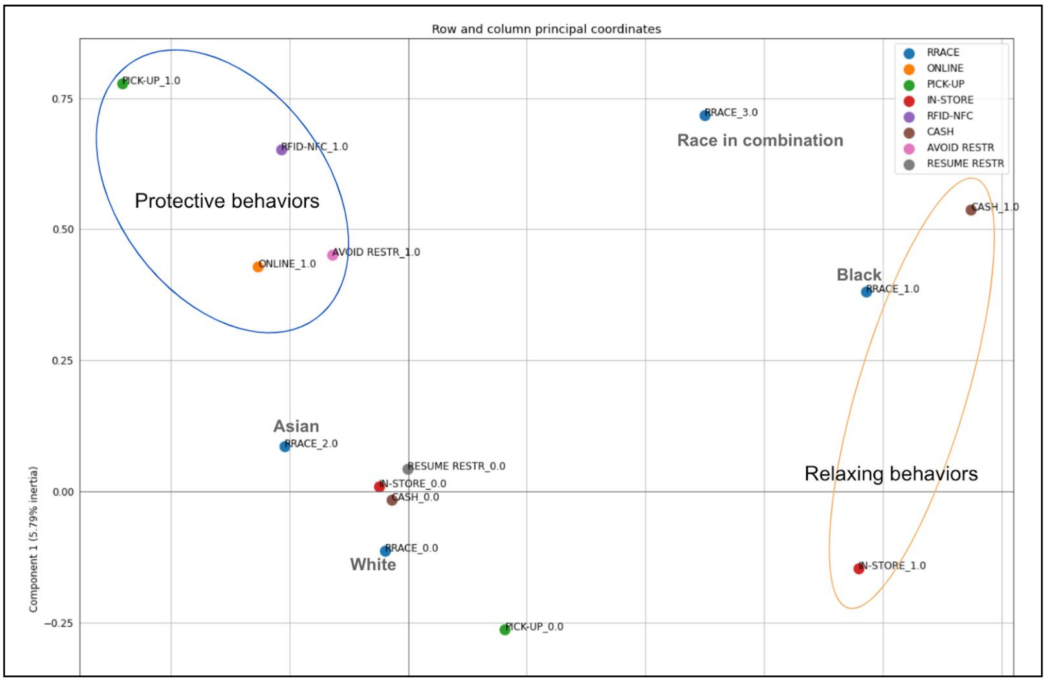
	Gender	Race	Education	Marital status	Adults + Kids	Income	Difficulty expenses
0	61% Women	90% White and Asian	23% Not college degree	74% Married	40% >= 1 child, 6% 1 adult, 70% 2 adults	74% above $150k	65% Not at all
1	69% Men	92% White and Asian	21% Not college degree	66% Married	70% no children, 15% 1 adult, 63% 2 adults	65% above $150k	74% Not at all
2	71% Women	86% White and Asian	22% Not college degree	68% Never Married	85% no children, 72% 1 adult	50% below $150k	67% Not at all
3	70% Women	38% black, asian, mix. 20% Hispanic	66% Not college degree	54% Never married	74% no children, 33% >=3 adults	88% below $150k	9% Not at all. 46% Some and very.

	Online	Pick-up	In-store	Contactless	Cash	Avoid Restaurants	Resume Restaurants
0	92% increase	37% increase	4% increase	73% increase	2% increase	88% increase	5% increase
1	23% increase	9% increase	7% increase	8% increase	2% increase	26% increase	8% increase
2	79% increase	23% increase	5% increase	22% increase	5% increase	75% increase	6% increase
3	38% increase	18% increase	12% increase	64% increase	5% increase	79% increase	4% increase

Table 3.1: Consumer Profiles in San Francisco using K-modes as clustering unsupervised technique

Then, MCA explains the relation between demographics and shopping variables based on the distance of a 2D projection of the variables. MCA is one alternative to PCA for dealing with categorical variables and allows us to extract new coordinates for the columns of our dataframe (the variables or features) and coordinates for the rows of the dataframe (feature vectors, i.e surveyed people). Two examples are displayed below (Figure 3.2). The variables included in the graphs are the shopping features and the demographics race and number of kids. As we can see, races in combination and black are closer to the CASH and IN-STORE variables, instead, race white and asian are near to ONLINE, CONTACTLESS, AVOID RESTR and PICK-UP. In the second plot, all the NUM_KIDS sub-categories are close to protective variables, but, in the absence of kids.

And finally, a visualization of the data points by cluster using the new MCA coordinates for the data points is performed using the projection of the data points in 2D (Figure 3.3)



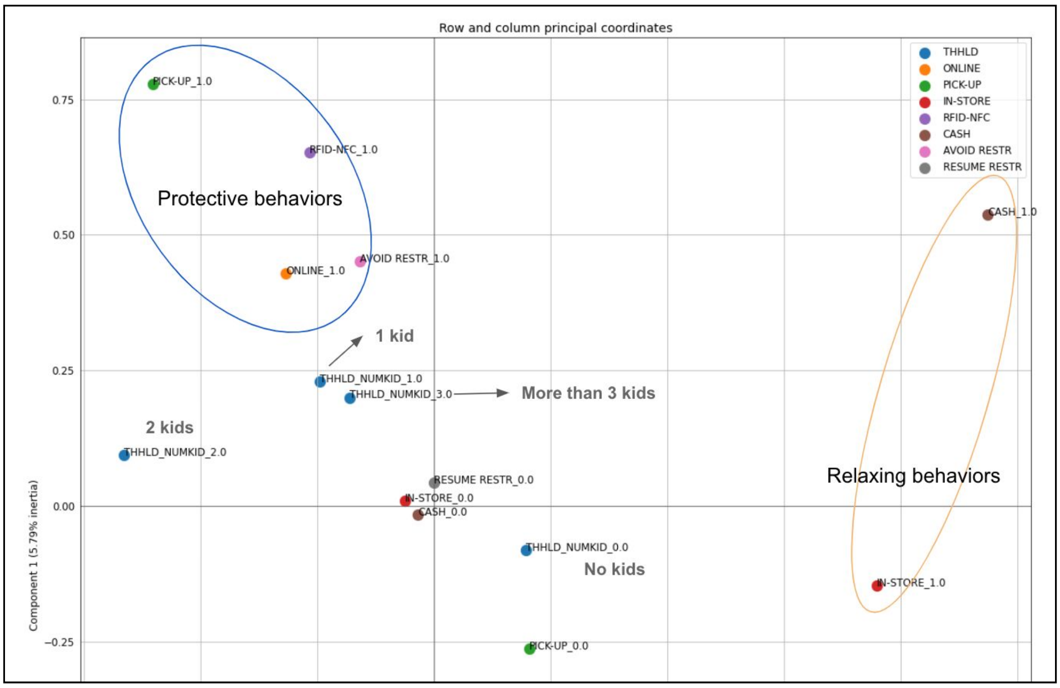


Figure 3.2: MCA dimensionality reduction for the sample population of San Francisco. Projection in 2D for shopping variables and race (up) and number of kids (down).

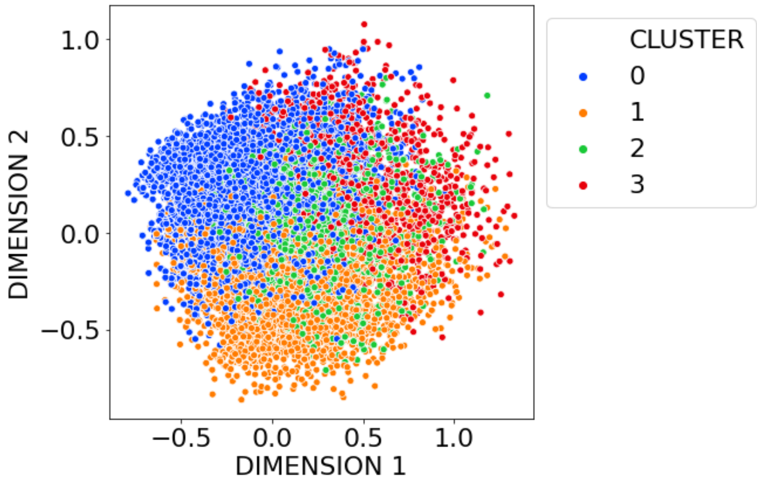


Figure 3.3: MCA 2D projection of the sample population using shopping and demographics variables to build an unsupervised clustering model(k-modes, with k=4)

3.1.2 Mobility Trends

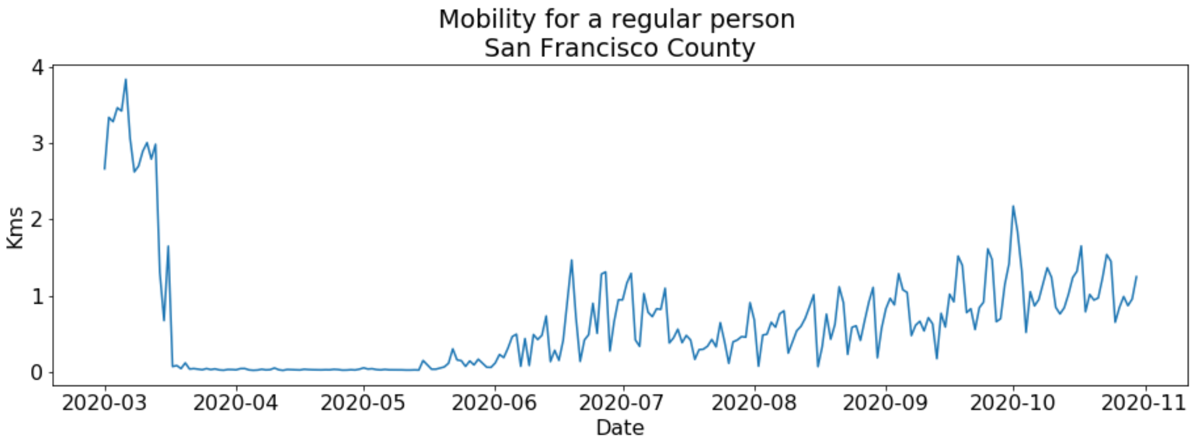


Figure 3.4: Mobility of a typical member of the community in The Bay Area

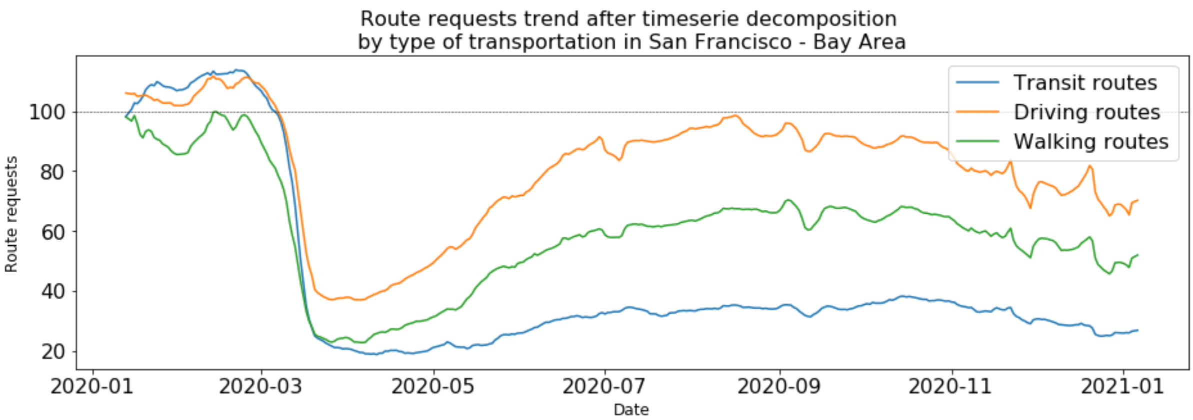


Figure 3.5: Trends of route requests in Apple devices by type of transportation

Shelter in place orders, fewer people physically going to work, temporarily closed and capacity restrictions in gyms, and multiple recommendations to stay at home and go out specifically for essentials have disrupted how much a typical member of the city moves in a regular day. The maximum average distance that people have moved over the pandemic is dramatically low in the San Francisco Bay Area, reaching between 20% and 50% of the baseline, which translates into **at most 1.5 km of distance** (Figure 3.7).

Besides the average distance for a regular person, commuting information can be explored directly from the route requests in Apple devices from the beginning of 2020 onwards. **All the commute types of transportations are under the baseline, but transit and walking are the most affected ones**. Apparently, the community is driving more than walking and using public transportation, in a city highly dependent on transit for commuting to work through BART train, Caltrain, Amtrak, MUNI train and transit, AC Transit (Figure 3.8)

On the other hand, the venues more affected with low foot traffic, according to Foursquare and Google Mobility data are **retail and recreation**, **transit stations, workplaces**, **airports**, **gyms**, **bars,** and **shopping malls**. Let's explore the map of the city to identify geographically the venues with more and less foot traffic and what are the features and current situation of restaurants there.

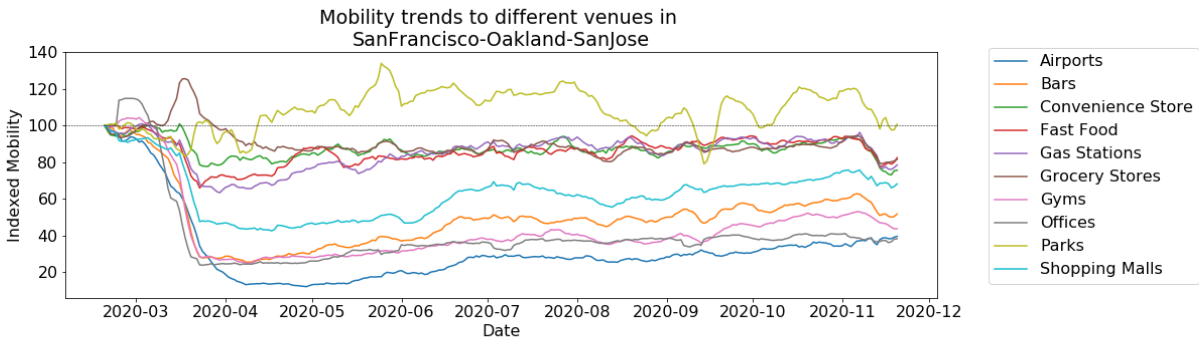


Figure 3.9: Foot traffic in parks was over the baseline during 2020. Shopping Malls, Gyms, Bars, Airports, and Offices indicate low foot traffic over the year.

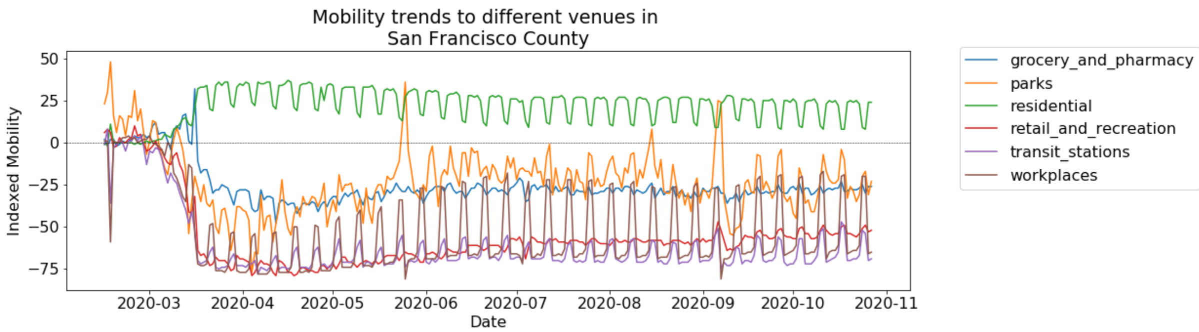


Figure 3.10: Foot traffic in residential areas was over the baseline during 2020. Transit, workplaces, and retail and recreation were the most affected venues.

3.2.3 Restaurants Scenario

The following analysis includes 50 businesses/every zip code of San Francisco county. The information of the restaurants was extracted using FUSION API Yelp, which allows a maximum of 50 results for an endpoint using the zip codes as keywords and words as *Restaurant.*

Appendix, Section 3.1, San Francisco Bay Area, Figures A14, A15 show a generalized residential plan and the main neighborhoods in the city. Both are used as references to identify parks and recreational areas, residential zones and how low, medium, and high dense they are, as well as commercial, industrial, mixed zones and rail transit systems to recognize transit stations and routes.

Popularity measured as the number of reviews and ranking

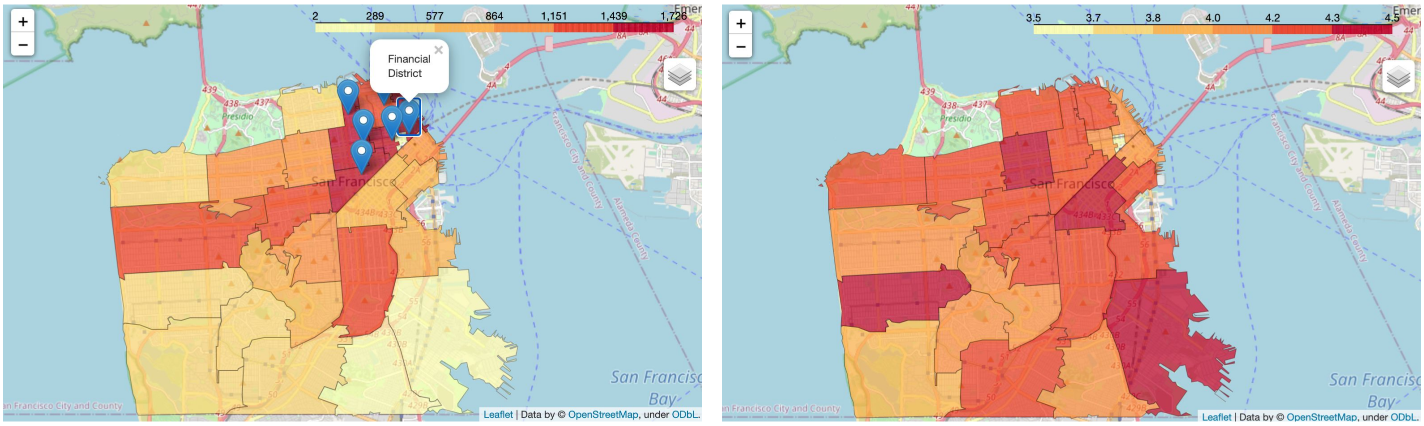


Figure 3.11: Average number of reviews by zip code (left) and average ranking by zip code (right)

The average number of reviews/zip code is shown in Figure 3.11 (left). The zones with higher average reviews are **Russian Hill, Nob Hill, Downtown, Chinatown,** and **Financial District**.

The zip codes with higher average ranking (Figure 3.11, right) are **Pacific Heights**, **Western Addition**, **Sunset District**, **Downtown**, **South of Market,** and **Hunter's Point.** Lake Merced has the lowest evaluation.

Ranking and reviews are interesting metrics to know the engagement of the customers with a business. Rankings alone don't bring enough information if we ignore the number of reviews used to calculate the ranking, and popular spots in tourist places usually have a lot more reviews than residential businesses. In this case, are considered as **hot spots** the zones with **high counts of reviews and peak rankings**, as the following Districts:

- Marina, Russian Hill, North Beach, Nob Hill, Pacific Heights, Downtown, Financial District, Chinatown, South of Market, and Golden Gate Park surroundings. From those spots, **the Financial District** and **South of Market** deserve attention because they have mostly offices, commercial and industrial buildings.

Prices and current transaction methodology

Yelp indicates as **$** a regular menu for a single person equivalent to less than $10; **$$** a menu for less than $30; **$$$,** a singular menu for less than $60 and **$$$$** whose menu exceeds that budget.

The zones with higher average prices are some of the most expensive zip codes in the city during 2020 as well (Pacific Heights, East of Richmond District, Twin Peaks zone, and Mission District). Russian Hill, Nob Hill, Fisherman's, Financial District, and South of Market reveal to be expensive spots too. Russian Hill and Nob Hill are residential areas (Figure 3.12).

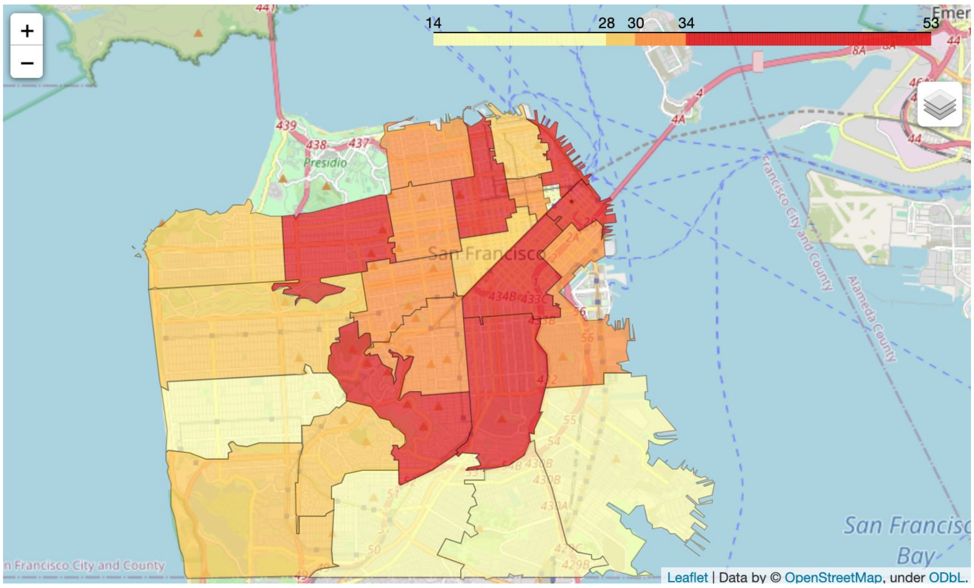


Figure 3.12: Average prices by zip code

Finally, the percentage of businesses offering delivery, pickup, or both services by zip code is displayed in Appendix, Section 3.1, San Francisco Bay Area, Figures A16, A17, A18. From A16, **Twin Peaks surrounding is the area offering more delivery services**, presumably because the zone is highly car-dependent, accessible by only one bus and it has equidistant proximity to every corner of the city. The businesses around the hills have the most strategic position in the city for delivery services. The next section generates recommendations based on the results of those Figures.

3.1.4 Discussion

These previous results are based on our dataset and don't make assumptions about the status of restaurants with missing transactional information. None of the businesses included in the analysis is expressly closed or listed as temporarily closed. All the conclusions are made under the assumption of reliable information, but most importantly, the thinking and analytical process can be replicated to understand the situation in other cities having similar or opposite scenarios, as we mentioned as Milestone 4 goal.

As we saw earlier, the businesses around the **Twin Peaks have a strategic position in the city for delivery services**, since they are equidistant from the entire county and can reach more customers and assure the same quality of service and timing for all of them, then restaurants in adjacent zones without delivery partnerships (**Noe Valley, The Castro, with an average price of $30-$34 and around 30% of businesses not listing delivery as purchase methodology) must incorporate delivery**.

Marina, Russian Hill, North Beach, Nob Hill, Pacific Heights, and Downtown are the zones with the highest density in the city, strongly residential. Since the mobility in the city indicates high foot traffic in the residential area, **it's very likely that the local community supports the businesses of the neighborhood,** but the support of the own community is not necessarily enough to keep businesses mostly dependent on the **presence of tourism** (as **Fisherman's Wharf District, with an average price of $34-$53 and around 30% of businesses not listing delivery as purchase methodology**).

According to the mobility data, parks and recreational zones are highly frequented. **The surroundings of Golden Gate Park** constitute hot **safe spots** because **they have high rankings and number of reviews**, but the **North and South of the park** have **completely different scenarios**. Businesses there are not popular enough and they are in residential but low-density areas. The average prices of businesses there are low, their transactional strategy is predominantly pickup, presumably because the local community is supporting their businesses. **They currently offer less than 50% of dual contactless modality and the rest as pickup exclusive**. Although residents of areas near the Golden Gate Park have access to Sunset and West Richmond District just walking an **average of 1.5kms**, **the use of delivery means reaching a larger audience.**

Lower and medium density residential areas (Lake Merced, Ingleside) **and commercial sectors** (Hunter's Point, Potrero Hill, South of Market) **are offering pick-up and delivery mixed** (in the first group, at least 50% of businesses are doing that and in the second group, between 50%-60%).

Excelsior has space to offer more delivery and pick-up since it is a medium-lower density residential zone.

Mission, as a mixed zone, with residences and commercial buildings could benefit from more pick up, because it's a zone of high transit and driving. **Haight** is a residential zone close to parks (more foot traffic) and it means that **pushing more pick-ups could help the local businesses there**. Finally, the **Financial District must increase their delivery options**. This is not a residential zone but **is merely 2 km walking from Chinatown, Nob Hill, and North Beach**.

3.2 Miami

Miami lies in stark contrast to San Francisco regarding COVID response. While San Francisco has maintained restrictions throughout the pandemic, Miami has had much looser restrictions. Ever since October of 2020, Miami restaurants have been allowed to operate at 100% capacity given that tables are spaced 6 feet apart. Before this point they operated at 50% capacity.

A factor in Miami's COVID response has been Governor Ron DeSantis. DeSantis had lifted all restrictions on dining in september of 2020 despite pushback from places like Miami. DeSantis is one of the most anti-restriction governors in the country. In contrast, Governor Gavin Newsom has been one of the most active governors in terms of placing restrictions.

Miami is also a different city than SF. The biggest industry driving GDP growth in Miami is Tourism and Hospitality. As covered earlier these industries have been hit the hardest by the pandemic. The biggest industry driving GDP growth in San Francisco continues to be the tech industry, which has largely gone remote. In theory, this means that SF can afford a lockdown more easily than Miami.

The sample population of Miami-Fort Lauderdale Pompano Beach indicates an increase in **online purchases** in the 54% of surveyed people, increase in the use of **contactless payment methods** in 92.35% and **avoidance of eating at restaurants** in 88.63%. It's also the metropolitan area with higher increment of in-store purchases in the Household Survey (10%).

3.2.1 Consumer preferences

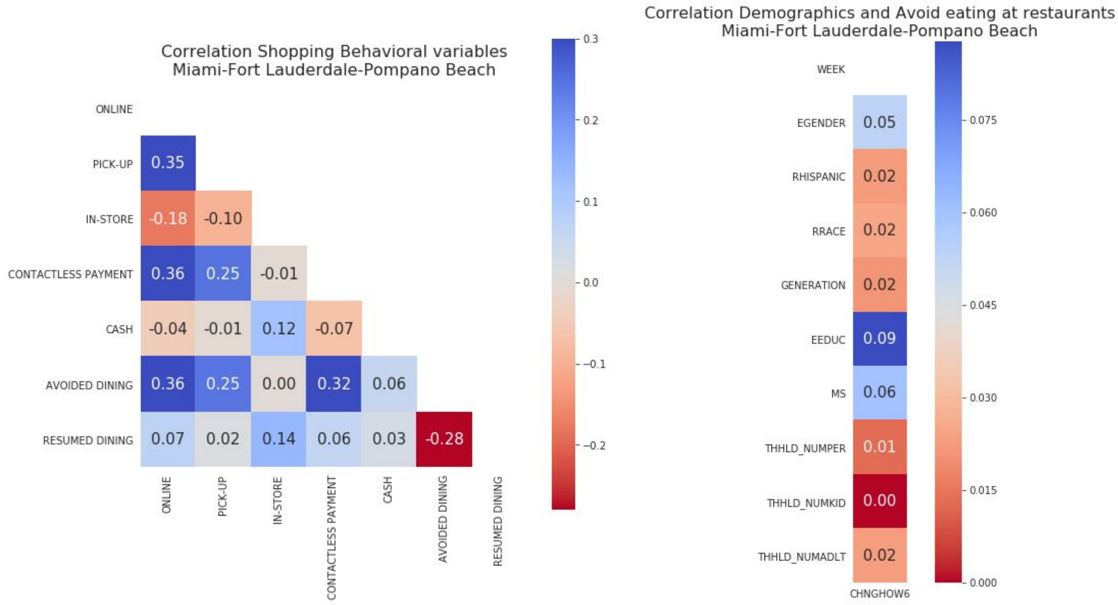


Figure 3.13: Lower triangular correlation matrix of shopping behavioral variables (left) and categorical association of demographic variables with avoidance of eating at restaurants (right).

According to Figure 3.13 (left), protective and relaxing variables are more correlated with each other than in San Francisco (in particular, the relationship between **ONLINE** and **avoidance of eating at restaurants** is stronger, as well as, between **IN-STORE** purchase and **resume eating at restaurants**). On the right, the correlation between **avoidance of restaurants and demographic variables** suggests that **educational attainment, marital status** and **gender play a role in the protective behavior** (no race and generation as in San Francisco). Diving into those variables, it was found a relationship between **avoidance of restaurants** and **women-higher educational attainment**. The analysis of shopping behavior by subcategories of demographics using lower triangular correlations indicates that:

- The variable most correlated to resumed eating at restaurants is white race.
- The variables most correlated to contactless payments are black race and 3, 5 and more adults in the household.
- Black people are more likely to complete in-store purchases. Low educational attainment is the second more related variable.
- People with higher educational attainment, married and 2 adults by household suggest the higher correlation with the increase of online purchases.

Next, we proceed to generate consumer profile clusters using K-modes and Silhouette analysis to determine the number of clusters . In this case, K=3 is the better choice.

The distribution of demographics and shopping variables by cluster can be found in Appendix, Section 3.2, Miami, Figure A19. Essentially, the main features by cluster are the following:

	Gender	Race	Education	Marital status	Adults + Kids	Income	Difficulty expenses
0	75% Women	36% Hispanic, 78% White, 14% Black	43% Not college degree	38% Married, 44% never married, divorced, separated	67% no kids, 43% 1 adult, 65% 4 adults	63% below $75k	38% Some and very.
1	60% Women	32 Hispanic,82% White, 11% Black	48% Not college degree	55% Married	70% no kid, 65% 2 adults	55% below $75k	54% Not at all
2	70% Men	33.4% Hispanic, 84% White, 9% Black	30% Not college degree	72% Married	66% no kids, 70% 2 adults	28% below $75k	47% Not at all

	Online	Pick-up	In-store	Contactless	Cash	Avoid Restaurants	Resume Restaurants
0	75% increase	28% increase	7% increase	39% increase	5% increase	87% increase	6% increase
1	13% increase	7% increase	12% increase	16% increase	4% increase	18% increase	13% increase
2	86% increase	36% increase	6% increase	79% increase	2% increase	85% increase	9% increase

Table 3.2: Consumer Profiles in Miami using K-modes as clustering unsupervised technique

From the MCA dimensionality reduction analysis, **low educational attainment**, **never married**, **separated and divorced variables** are closer to the **increase of CASH**. Again, the **absence of kids is closer to relaxing behaviors**, as well as **low income levels.**

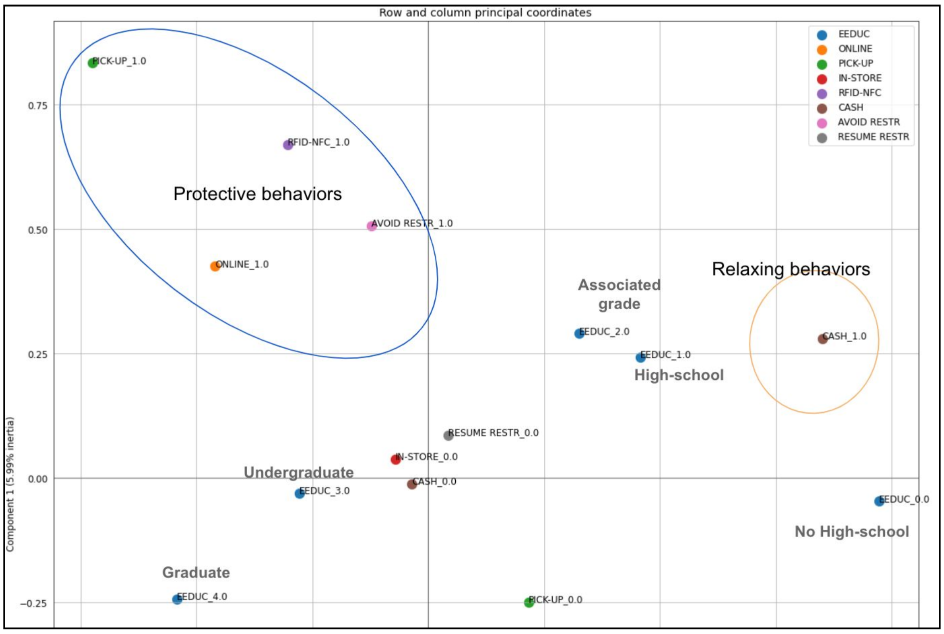


Figure 3.14: MCA dimensionality reduction for the sample population of Miami. Projection in 2D for shopping variables and educational attainment

Finally, a visualization of the data points by cluster using the new MCA coordinates for the data points in Miami is performed using the projection of the data points in 2D.

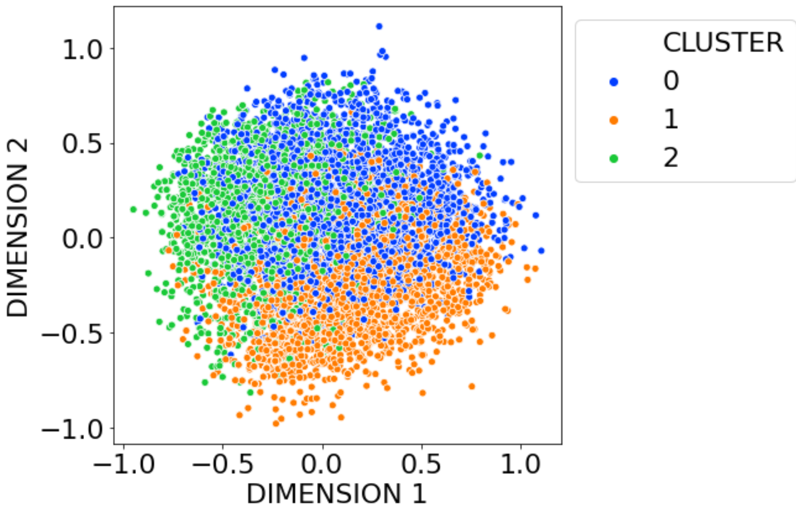


Figure 3.15: MCA 2D projection of the sample population of Miami using shopping and demographics variables to build an unsupervised clustering model(k-modes, with k=3)

3.2.2 Mobility Trends

The Descartes chart (Figure 3.16), reveals that mobility for the average person plummeted during the beginning of the lockdowns. At the baseline, the average person travelled 8km but this went down to around 1km shortly after the first shelter in place orders in March. The highest average mobility after the beginning of the COVID crisis is seen in October 2020 where mobility reached almost **7km** or about **80% of baseline**.

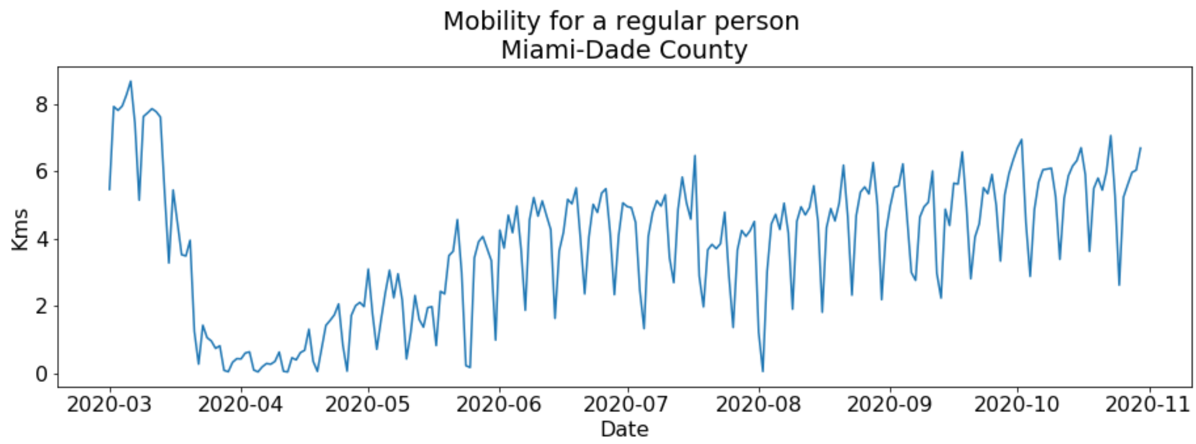


Figure 3.16: Mobility of a typical member of the community in Miami Dade County

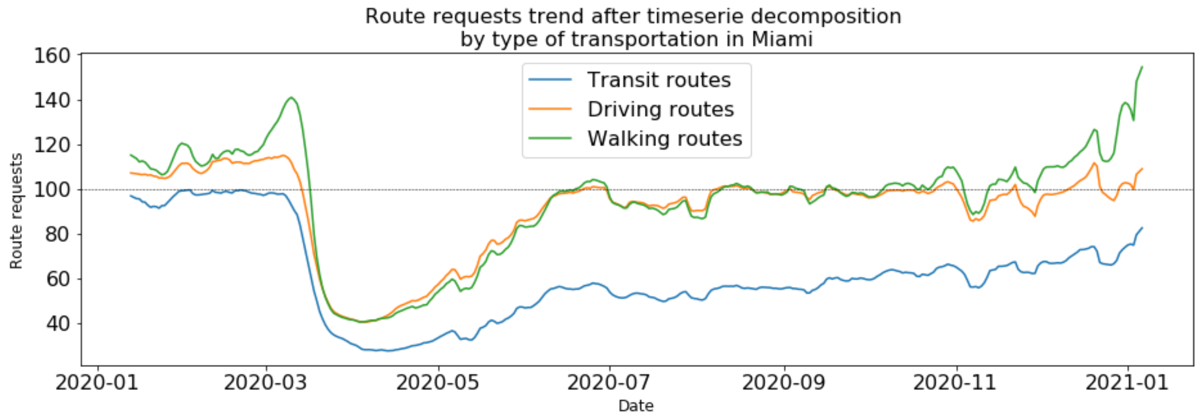


Figure 3.17: Trends of route requests in Apple devices by type of transportation

The different types of route requests seen in Apple mobility highlights the following: as expected, **public transportation requests went down severely during the crisis** and always have stayed well below baseline. **Walking and driving also decreased** but went back up to baseline in July 2020. **Driving remains at baseline for the rest of the pandemic** but **walking actually increases above baseline towards the end of 2020,** even beating out the pre-covid walking peak.

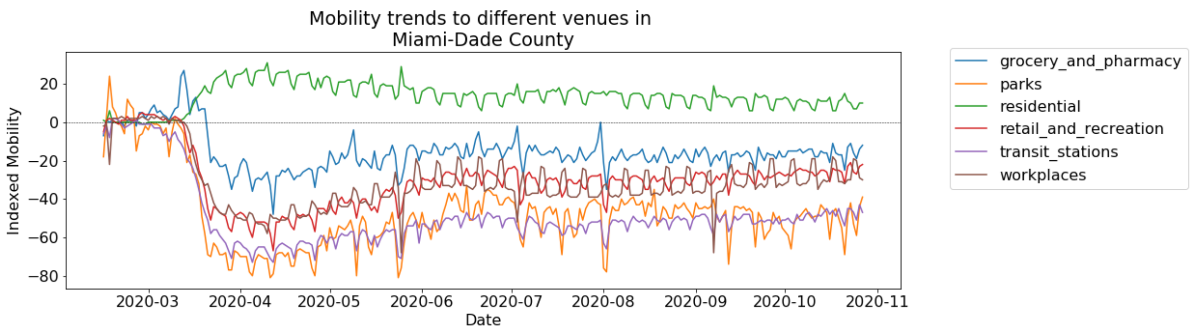


Figure 3.18: Foot traffic in parks was over the baseline during 2020 in Miami. Parks, transit stations, retail and workplaces indicate low foot traffic over the year.

Transit stations, parks, retail & recreation, and workplaces have been the most severely affected venues according to Google Mobility Data (Figure 3.18). Grocery stores and pharmacies have had less disruption, possibly due to their necessity. Residential venues have climbed in traffic but that is to be expected with the shelter in place measures.

3.2.3 Restaurants Scenario

According to Appendix, Section 3.2, Miami, Figure A20, the zones more densely populated in Miami Dade County are the city of Miami (**Downtown, Design District, East Little Havana**), **Hialeah, Palm Spring North**, **Sunny Beach**, **Miami Beach**, **Fontainebleau** and **Flagam**i. Analyzing restaurants popularity in each zip code as the mean of reviews by restaurants, calculating weighted rankings (using number of reviews and rankings) and finally, extracting the percentage of businesses by zip code purchasing by delivery exclusively, pickup exclusively or both, we get the following results:

- More than 55% of restaurants in zip codes around the Miami International Airport between percentile 50 and 85 in weighted rankings and popularity (as number of reviews) are offering **only delivery services**, which makes sense because **they are not densely populated regions and the airport is reporting low foot traffic since the pandemic**.
- Shopping zones around Overtown and East Little Havana have medium dense populations and low foot traffic. They are already offering almost 60% delivery only and 40% delivery and pick-up.
- South Miami and Westchester are low density zones. More than 65% of businesses in those zones are already offering only delivery, which is the **best option in low traffic areas with low population.**
- An interesting observation in this case is the high rates of delivery exclusive or delivery and pick up that businesses are offering. Only 2 zip codes offer more than 30% pick-up exclusively and they are close to medium density residential zones to justify the choice. All the zip codes so far of dense areas offer a mix of both modalities or delivery exclusive. Instead of recommending more partnerships with delivery companies, as we discuss in the previous case study in this scenario, the question should be: Which businesses do not require so much delivery because they could benefit from pickup services due to the density of population in their areas?
    - North Miami and Allapattah: 100% of businesses offering delivery. The zones are medium densely populated and can benefit from pick-up, avoiding being charged for delivery services.

Frontend Design

The culmination of the project consists of an analytical Dashboard, which combines consumer preferences, mobility patterns, and the current situation of restaurants for a particular location. As part of the in-depth analysis and the presentation of the project, our team is going to expose how to use the dashboard to get insights for restaurants in a specific city.

Business Metrics

1. **Consumer preferences** metrics: shopping variables allows us to understand **how likely is the population to prefer delivery food services, pick-up services, and resume indoor and outdoor dining.**
    a. Purchase preferences: Percentage of the population doing more shopping through one of the following options: online, pick-up, in-store.
    b. Payment methods: Percentage of the population using more contactless payment methods and cash.
    c. Likelihood to come back to restaurants: Percentage of people reporting resumed/avoided eating at restaurants.
2. Consumer Clustering: Use of k-Means to perform clustering of the surveyed population to identify groups of people based on demographics and shopping behaviors. This is useful to determine **which groups are more or less likely to resume indoor and outdoor dining.**
3. Consumer Heatmaps to show dependency between variables. The values in the heat maps represent the statistical value for a chi-squared statistical test performed between a demographic and a shopping behavioral variable. Values highlighted don't reject the null hypothesis about dependency with a 95% of confidence.

The source used to get those metrics is the Household Pulse Survey 2020, from week 13th to 20th of the survey, which means August 19th to December 7th, 2020.

4. **Mobility patterns**
    a. Comparison of foot traffic in different venues (index foot traffic, using a **baseline day.** A baseline day represents a *normal* value for that day of the week. The baseline day is the median value from the period Jan 3 – Feb 6, 2020). This data provides us **information about the venues with more and less traffic** and helps us to **identify areas more disrupted into the cities**. Less foot traffic in the gyms, offices, airports, and commercial areas leaves the surrounding restaurants at risk.

    b. Comparison of route requests in Apple devices by type of transportation (index route requests related to a baseline volume, January 13th, 2020). This is useful to identify the type of transportation more disrupted. More people driving or walking represent opportunities for pick-up services; less use of public transportation leaves the nearest businesses to bus and train stations at risk.

    c. Index mobility for a regular member of the community, based on a baseline defined as the median of max distance mobility measured during 2020-02-17 to 2020-03-07. This trend is extremely useful to **determine how likely people are to use pick-up services**. We use this information to determine the maximum distance that people are willing to move from their residences to buy food through pick-up. Lower indexes represent opportunities for delivery.

The sources used in this section are The Foursquare Mobility Data, Google Mobility Data, Descartes Lab Data, and Apple Mobility Data.

5. **Restaurants:** The last section includes a map of the city with restaurants by zip code. Every restaurant provides information about prices, type of food, purchase methodology, rating, and the number of reviews.

   The source used in this section is Yelp Data through the use of the FUSION API.

Operational and analytical data presented

The dashboard is currently available in the following [link]() as a preliminary version. The first page contains information about the in-depth analysis of the restaurant scenario in San Francisco and the second page replicates the same analysis for the city of Miami. The conclusions and recommendations of each case, which are included in the Report and as part of the Final Presentation, are not presented in the dashboard, since this is a visualization tool provided to the client to understand the consumer, mobility, and current context of the restaurants in a particular city.

Interactive data visualizations

The user can interact with the dashboard selecting shopping metrics (Figure 1), mobility foot traffic trends by source (google and foursquare data) as exposed in Figure 2, and filtering restaurants by type of purchases currently offered, prices, zip-code, and type of food (Figure 3). This last variable, type of food, requires *keywords* as the following: *Asian, Mediterranean, Indian, Peruvian, American*. Finally, Figure 4 displays the Dashboard for SF.

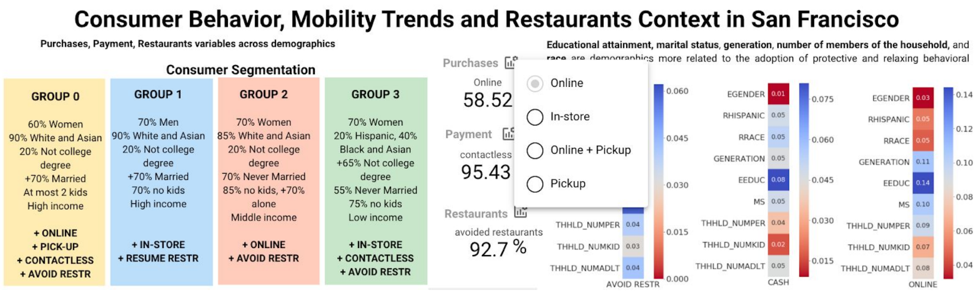


Figure 1: Interactive use of Section 1 of the dashboard. The user can select between purchases, payment, and restaurant preferences of people surveyed by the Household Pulse Survey 2020

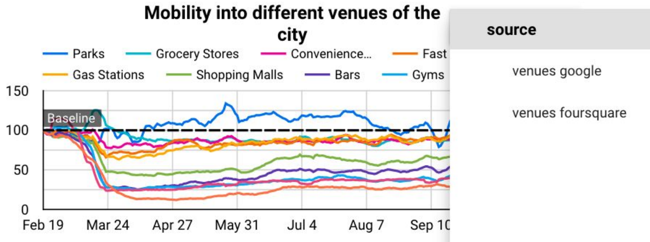


Figure 2: Interactive use of Section 2 of the dashboard: The user can select the tracking of different venues based on their baseline choosing google or foursquare classification of venues.

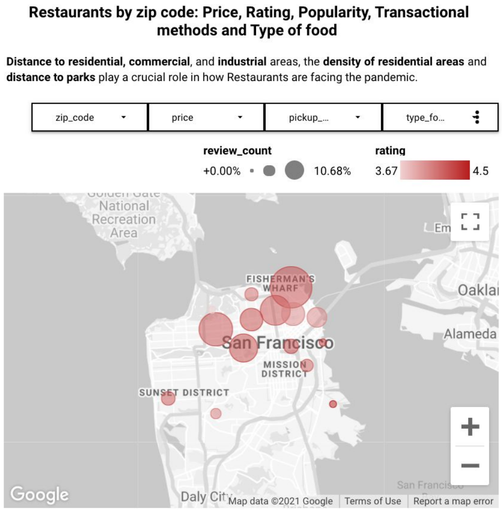


Figure 3: Interactive use of Section 3 of the dashboard: The user can select specific zip codes, prices, and purchase modalities offered by the restaurants, as well as types of food.

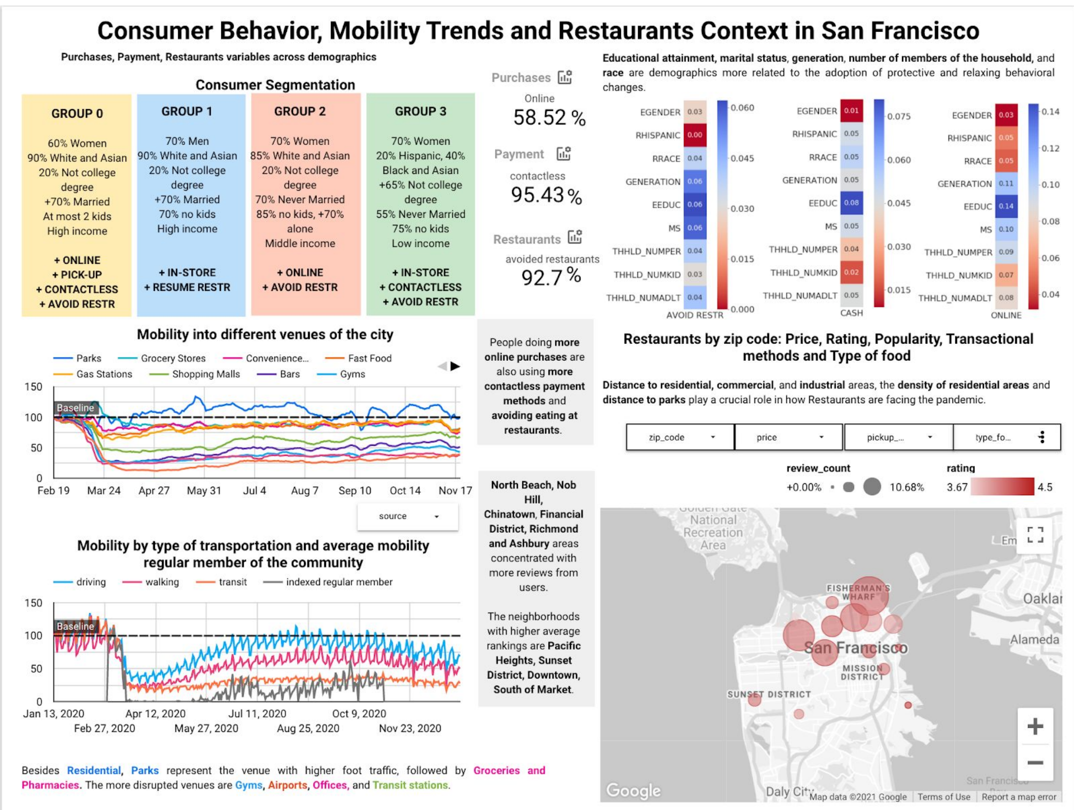


Figure 4: Dashboard of San Francisco

Further Research

The case studies developed in the dashboard can be extended for more cities and specific locations across the country. As a further step, we could pick one representative city in the Northeast and Midwest regions and compare the scenarios and recommendations with the cities from the West and South included in this report, as well as, pick more cities in every region to extrapolate representative recommendations by region. Since our datasets include shopping, mobility and restaurant information from other locations, we can build automated reports using, for example, the DataDog API.

From the restaurant data, more data of restaurants in every location, including temporarily closed businesses, traffic into the websites and orders through delivery apps versus businesses websites, will provide meaningful insights to discover how much restaurants need to invest in technology on their own. Additionally, the inclusion of all the restaurants from the cities analyzed would bring a more precise picture of the impact that lockdowns have had in the industry locally. Foot traffic by category of restaurants or by zip codes and shopping behaviors split into more specific sub-sections can also bring invaluable details to improve our clusters and likelihood of foot traffic into the city.

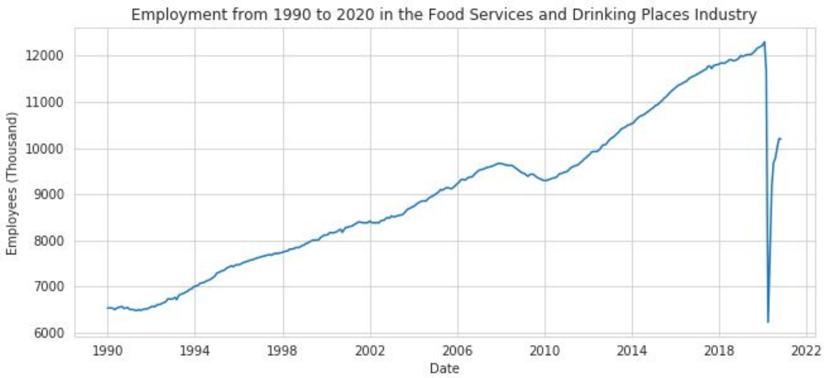
Appendix

Section 1: Economic Perspective

Economic Rows	June-2018	September-2018	December-2018	March-2019	June-2019	September-2019	December-2019	March-2020	June-2020	September-2020
Personal consumption expenditures (PCE)	1.34%	1.05%	0.75%	0.60%	1.55%	1.02%	0.78%	-1.45%	-9.96%	9.90%
Goods	1.23%	0.60%	0.31%	0.16%	2.32%	0.79%	0.19%	-0.21%	-4.20%	11.57%
Durable goods	1.42%	0.41%	0.30%	0.09%	2.78%	1.09%	0.08%	-3.71%	-1.21%	18.63%
Motor vehicles and parts	1.06%	0.46%	0.15%	-3.40%	3.23 %	0.15%	0.48%	-8.25%	0.00%	20.80%
Furnishings and durable household equipment	1.92%	0.49%	-0.20%	1.27%	2.20%	1.23%	-0.25%	-0.47%	-2.00%	15.85%
Recreational goods and vehicles	1.30%	1.08%	1.02%	3.23%	3.25%	2.04%	0.14%	0.05%	6.75%	11.47%
Other durable goods	1.73%	-1.01%	0.05%	0.79%	1.79%	1.26%	-0.40%	-5.60%	-19.39%	38.12%
Nondurable goods	1.14%	0.69%	0.31%	0.19%	2.09%	0.64%	0.25%	1.61%	-5.67%	7.95%
Food and beverages purchased for off-premises consumption	0.50%	0.45%	0.44%	0.61%	1.16%	1.15%	-0.27%	7.80%	2.17%	0.77%
Clothing and footwear	1.88%	0.13%	0.63%	0.20%	1.35%	0.30%	-0.17%	-9.48%	-21.47%	31.65%
Gasoline and other energy goods	1.72%	1.95%	-2.55%	-7.31%	7.15%	-3.02%	1.76%	-8.81%	-39.34%	31.65%
Other nondurable goods	1.25%	0.71%	0.97%	2.07%	1.76%	1.36%	0.41%	2.94%	0.26%	5.50%
Services	1.39%	1.26%	0.95%	0.80%	1.20%	1.12%	1.04%	-2.00%	-12.58%	9.07%
Household consumption expenditures (for services)	1.34%	1.29%	0.72%	1.10%	1.24%	1.10%	1.15%	-2.95%	-13.91%	11.03%
Housing and utilities	1.23%	0.81%	1.25%	1.02%	1.13%	1.12%	0.70%	0.72%	1.60%	0.69%
Health care	1.12%	1.62%	0.01%	1.49%	1.43%	0.69%	1.76%	-3.87%	-16.75%	19.26%
Transportation services	-0.02%	-0.06%	0.97%	0.26%	1.91%	1.87%	0.90%	-8.13%	-36.74%	24.16%

Recreation services	0.34%	1.69%	0.35%	0.55%	1.43%	-0.02%	1.90%	-9.16%	-45.94%	38.99%
Food services and accommodations	1.55%	1.88%	-0.26%	0.72%	1.93%	1.16%	0.20%	-8.41%	-33.51%	33.86%
Financial services and insurance	1.15%	1.60%	1.38%	0.73%	1.48%	1.34%	1.27%	0.12%	-1.57%	2.64%
Other services	3.15%	1.29%	1.22%	1.81%	0.00%	1.84%	1.33%	-2.59%	-17.86%	7.31%
Final consumption expenditures of nonprofit institutions serving households (NPISHs)1	2.38%	0.46%	5.78%	-5.33%	0.18%	1.58%	-1.35%	18.96%	11.72%	-18.45%
Gross output of nonprofit institutions2	2.21%	1.60%	0.77%	0.85%	0.60%	0.64%	0.97%	0.18%	-5.09%	2.35%
Less: Receipts from sales of goods and services by nonprofit institutions3	2.14%	2.03%	-1.09%	3.29%	0.76%	0.30%	1.82%	-6.46%	-12.63%	14.31%

Table A1: PCE quarter-by-quarter basis



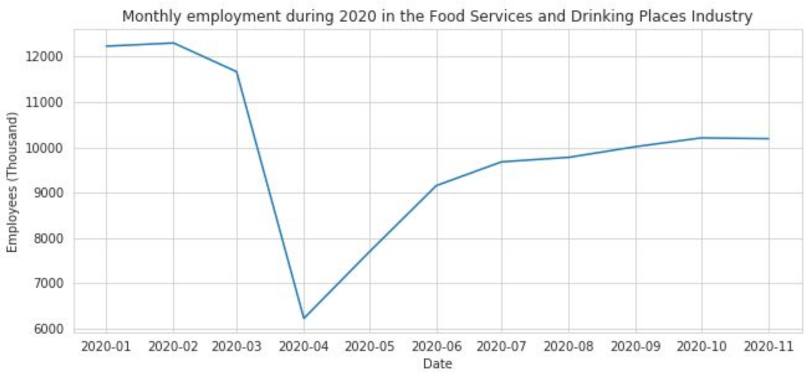


Figure A1: Employment in the restaurant industry from 1990 to 2020: On the upper plot, we explore the number of employees in the industry (thousand) from 1990 to 2020. On the lower plot, the same curve during 2020. Among the 20.5 million U.S. jobs lost in April, about 5.5 million of them were in the restaurant industry. The huge loss has erased about three decades' worth of restaurant and bar jobs, with employment levels in the industry back to where they were in the late '80s.

Section 2: Consumer Perspective

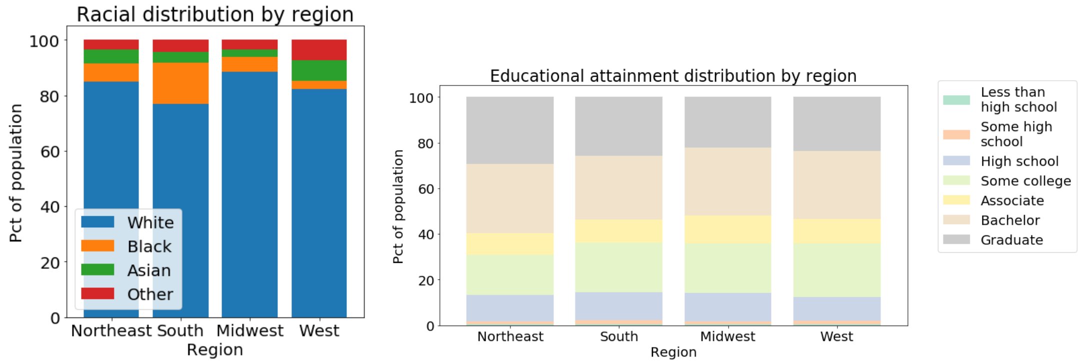


Figure A2: Percentage of the sample population by region and race (on the left) and population by region and educational attainment (on the right).

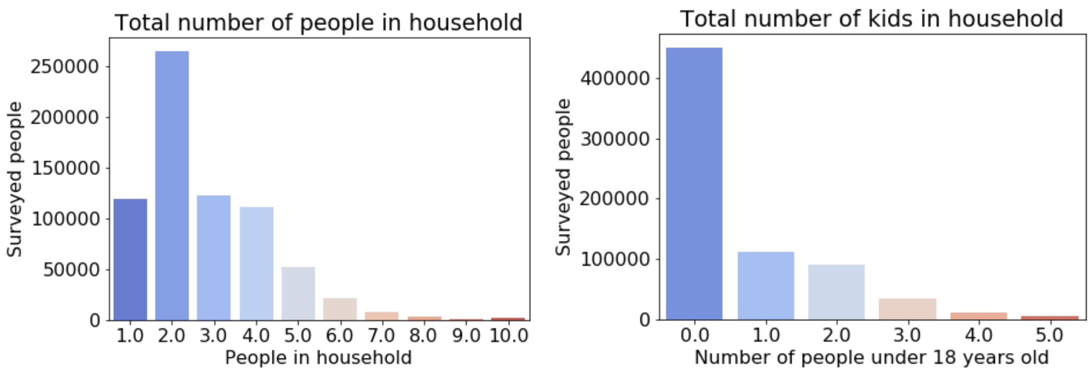


Figure A3: Total number of kids and members of the household across the Household Survey during weeks 13th-20th.

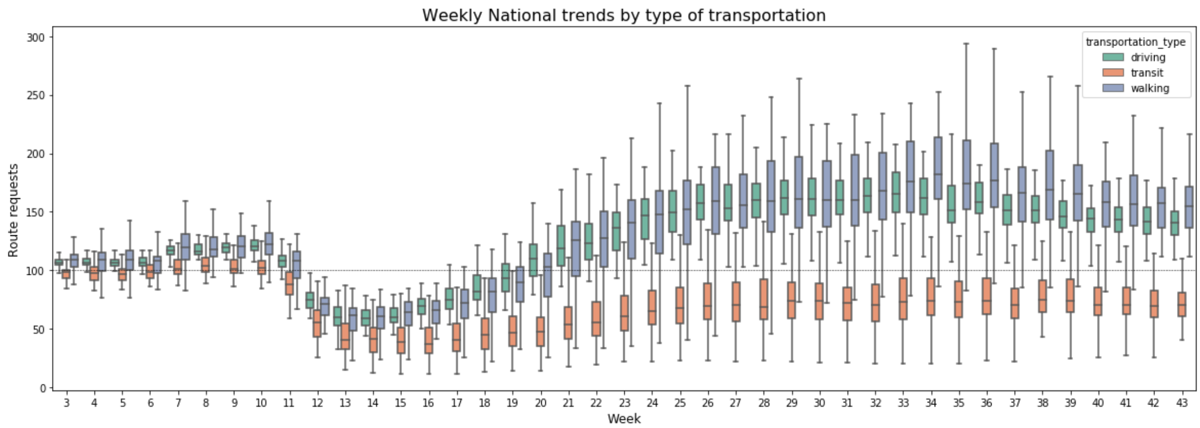


Figure A4: Weekly directions requests across the country during the year: Between the weeks 11th and 13th governors in different states announced statewide orders to stay at home for the non-essential workforce (California, New York, New Jersey, Ohio, Louisiana, Massachusetts). It's interesting to note that between weeks 11 and 19, the mean of route requests for driving and walking are under the baseline across the country. From there, the mean of directions requests is over the median, and the deviation of the weekly average by state increases. The increment of the percentiles 50 and 75 of the weekly requests answers to seasonality (summer vacations) and the variability into every boxplot can be explained for the independent management that every state did relate to the stay-at-home restrictions.

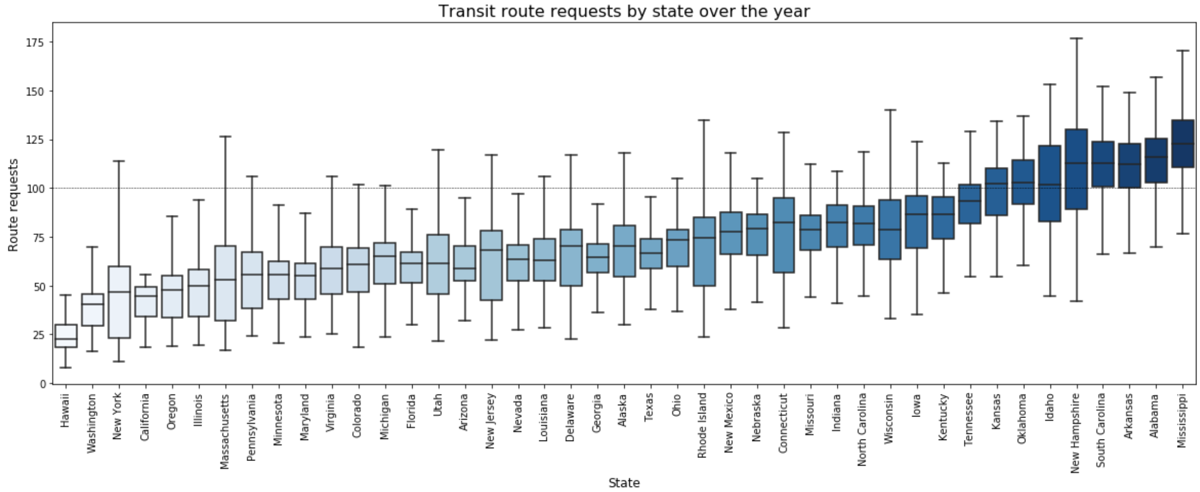


Figure A5: Transit requests by state during the year

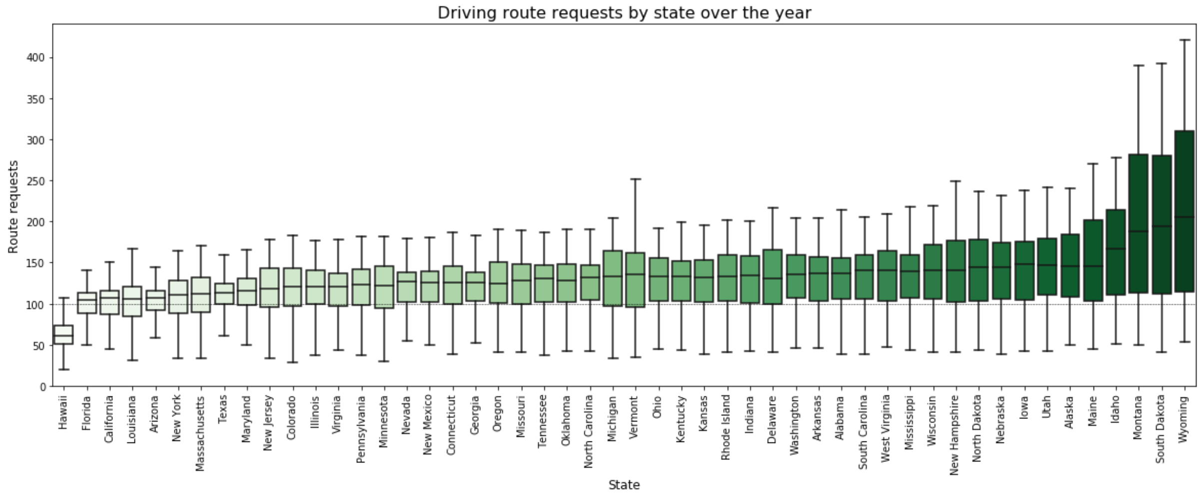


Figure A6: Driving requests by state during the year

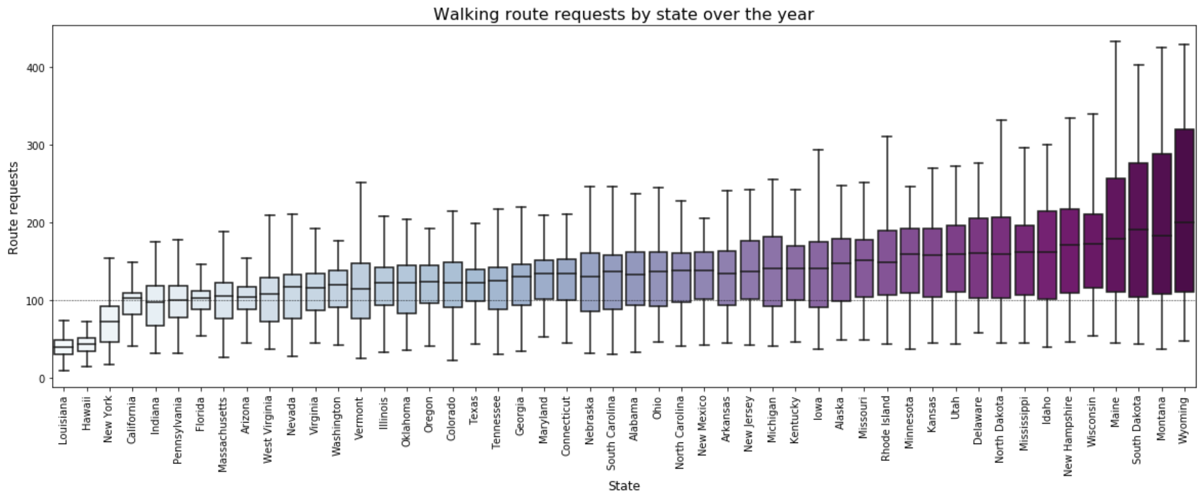


Figure A7: Walking requests by state during the year

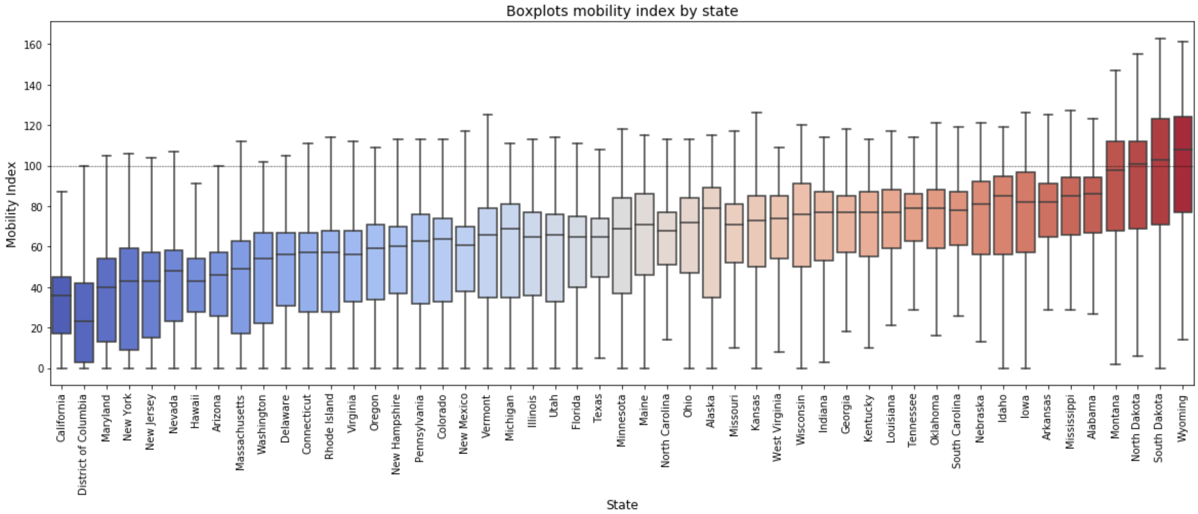


Figure A8: Indexed mobility changes by state

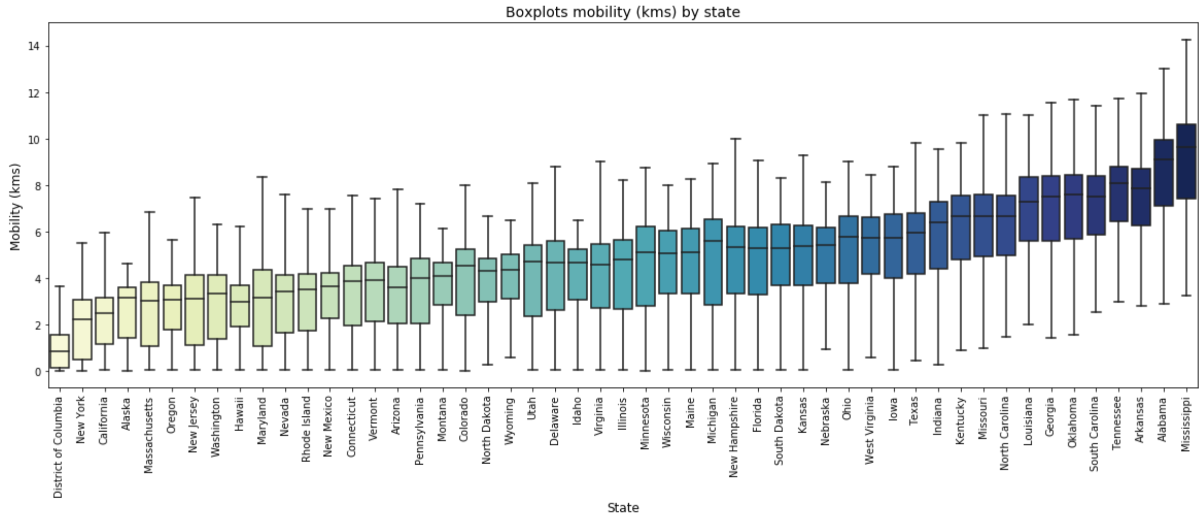


Figure A9: Mobility changes (kms) by state

Section 3: In-Depth EDA

3.1 San Francisco Bay Area

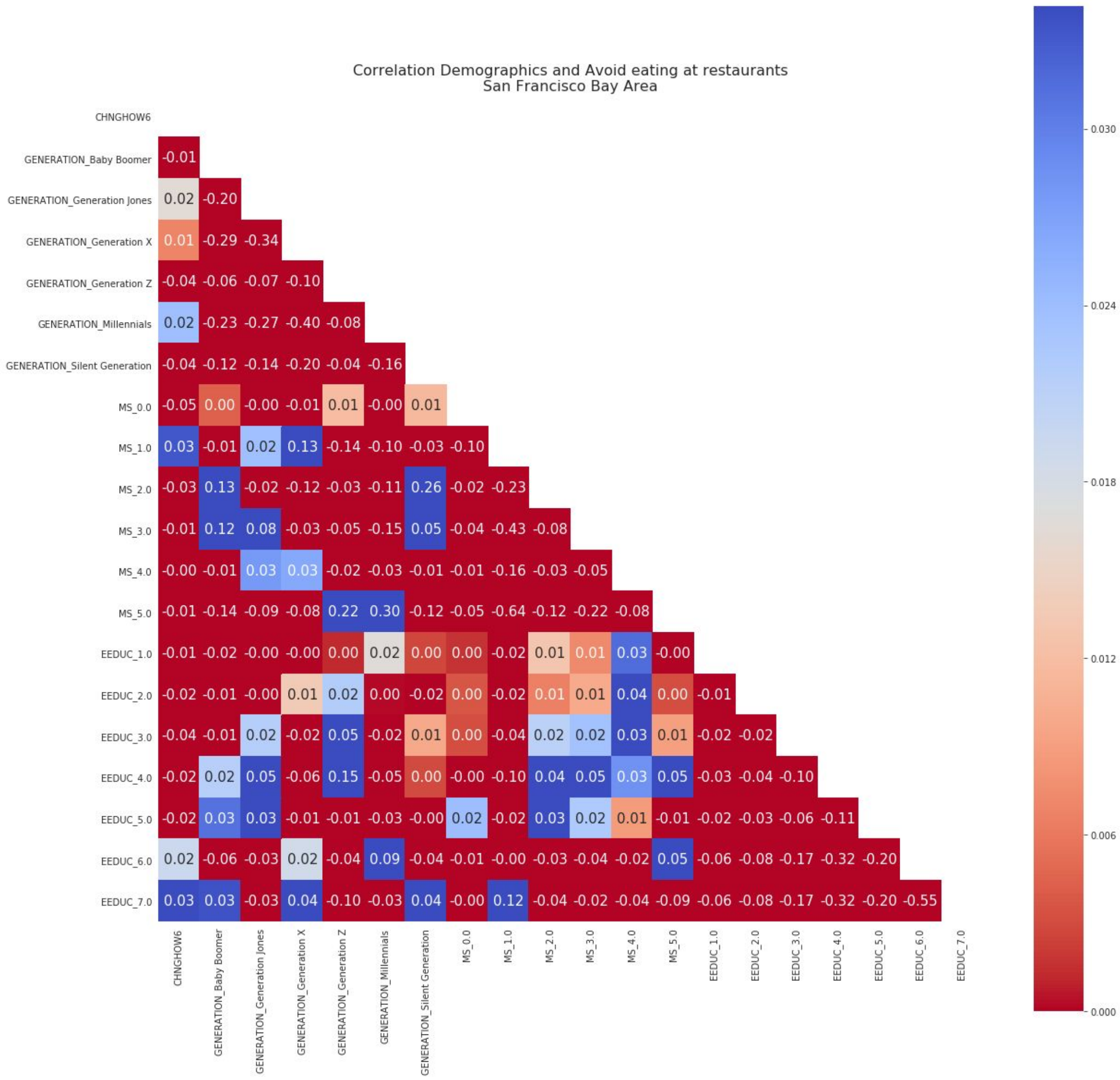


Figure A10: Lower triangular correlation matrix of **avoidance of eating at restaurants** and the demographic categorical variables in their one-hot encoding version: the analyzed shopping variable is related to marital status 1 (married), high educational attainment (Master and Ph.D.), and Generation X, Jones, and Millennials

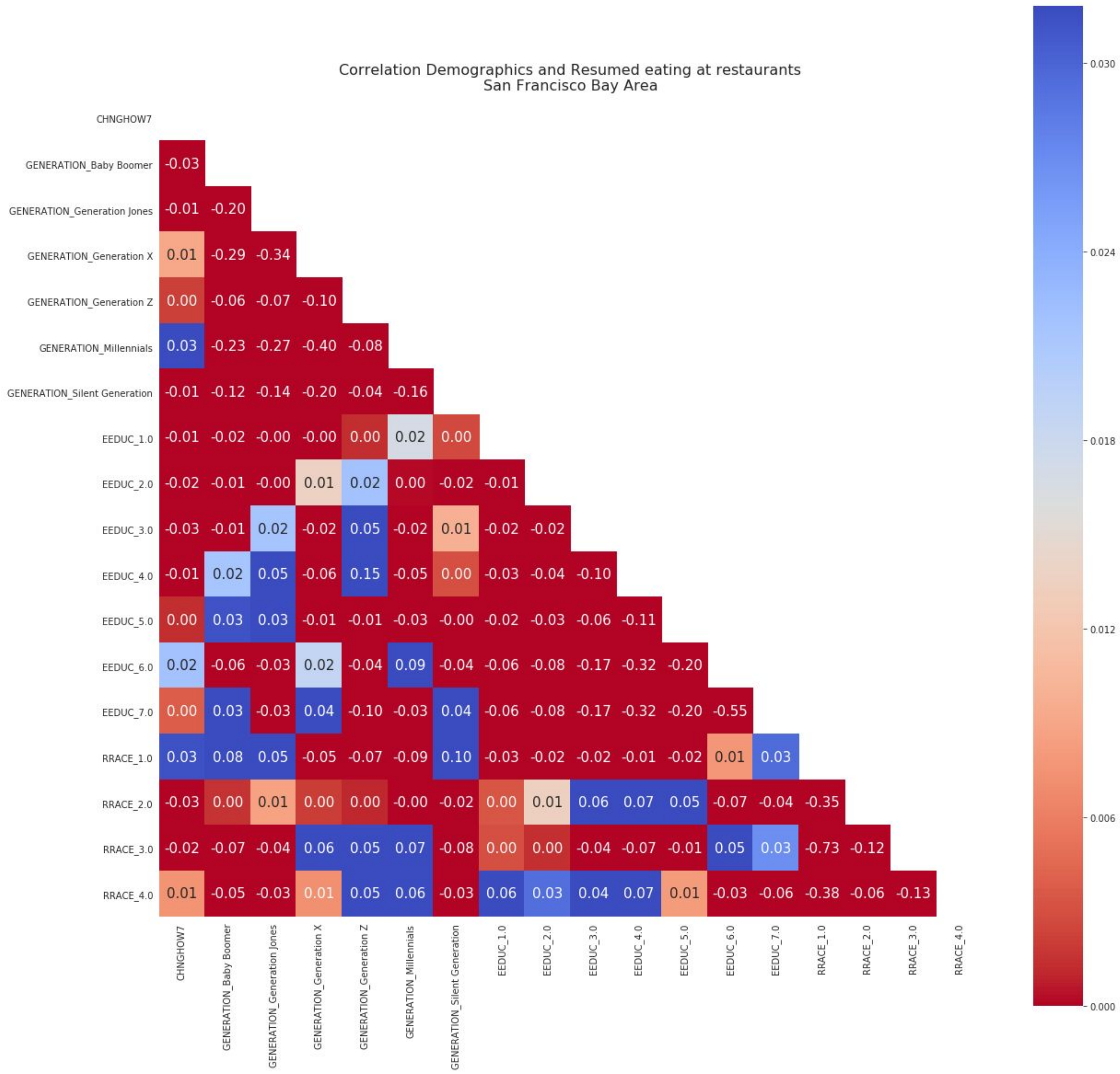


Figure 3.3: Lower triangular correlation matrix of **resume eating at restaurants** and the demographic categorical variables in their one-hot encoding version: the analyzed shopping variable is related to race 1 (white), high educational attainment (college degree), and Millennials

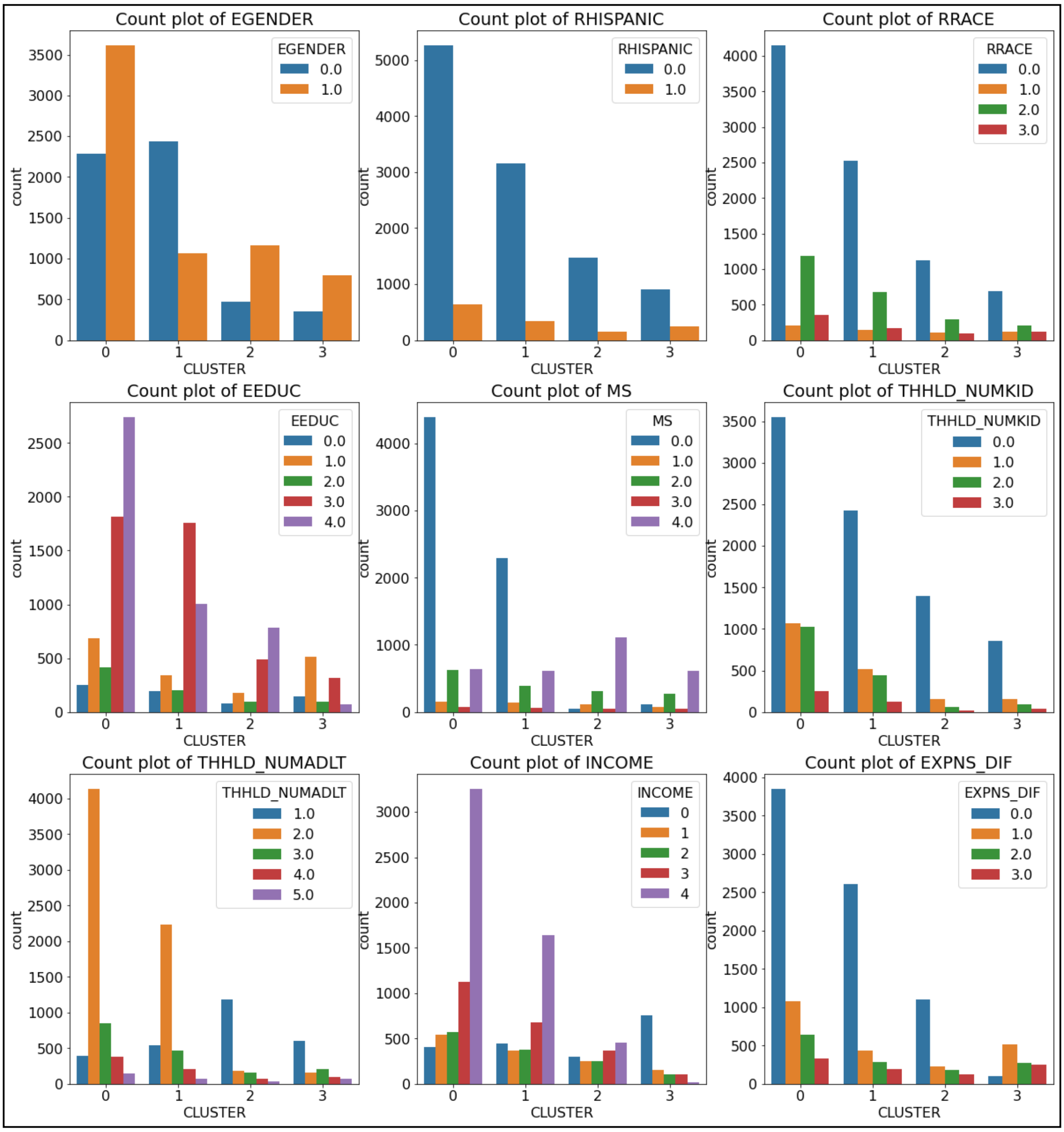


Figure A12: The distribution of demographics by cluster in San Francisco Bay Area

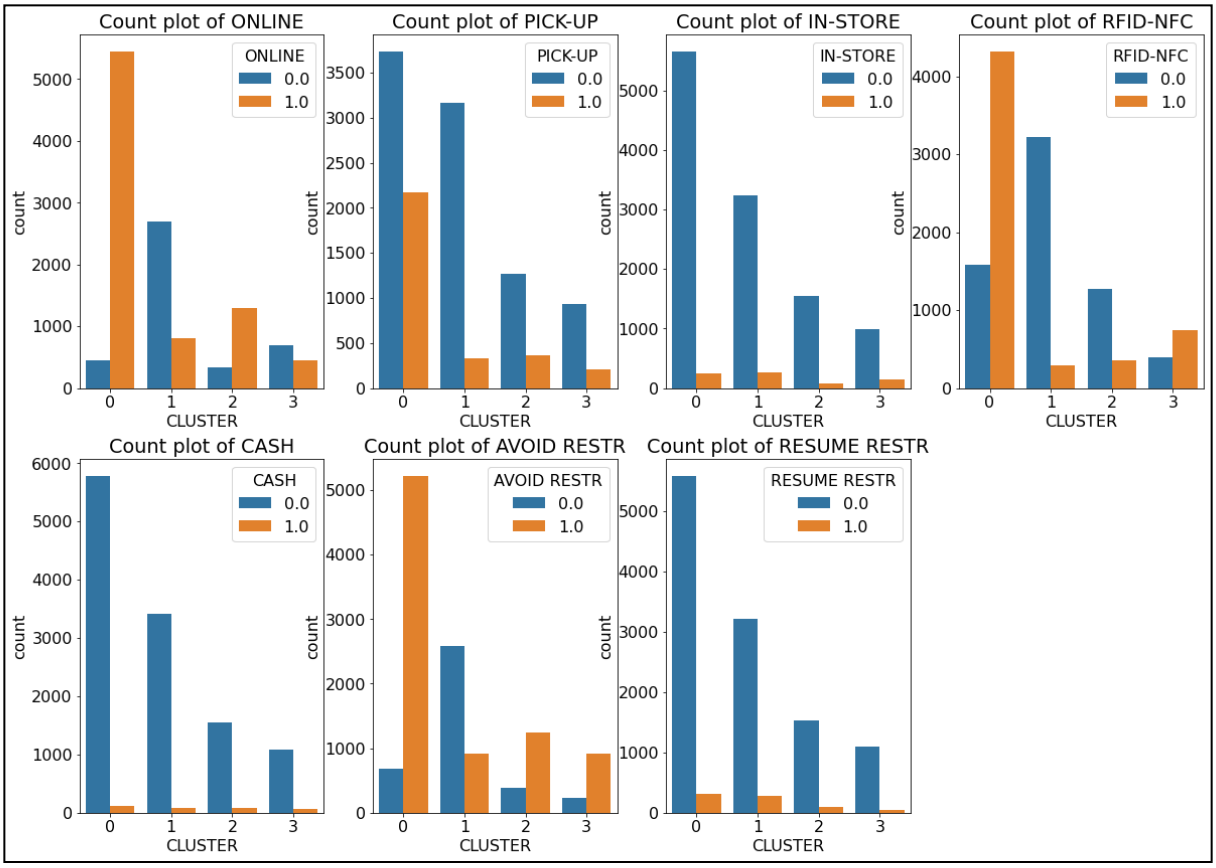


Figure A13: The distribution of shopping preferences by cluster in San Francisco Bay Area

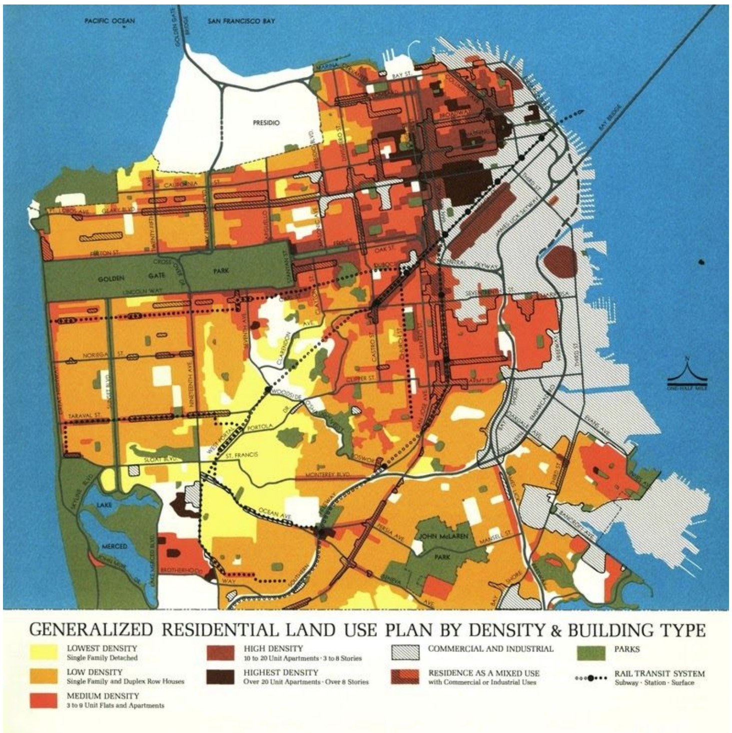


Figure A14: Referential distribution of land in San Francisco county

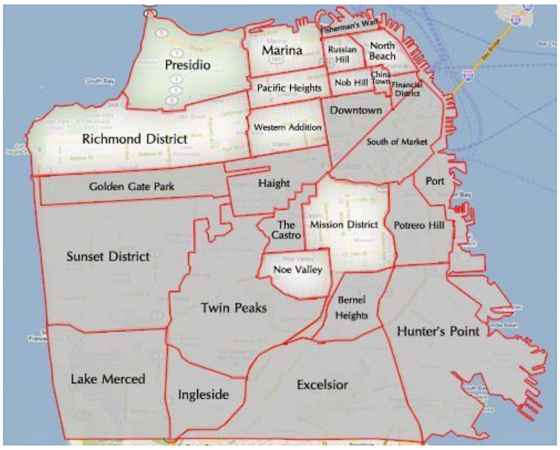


Figure A15: San Francisco county neighborhood divisions

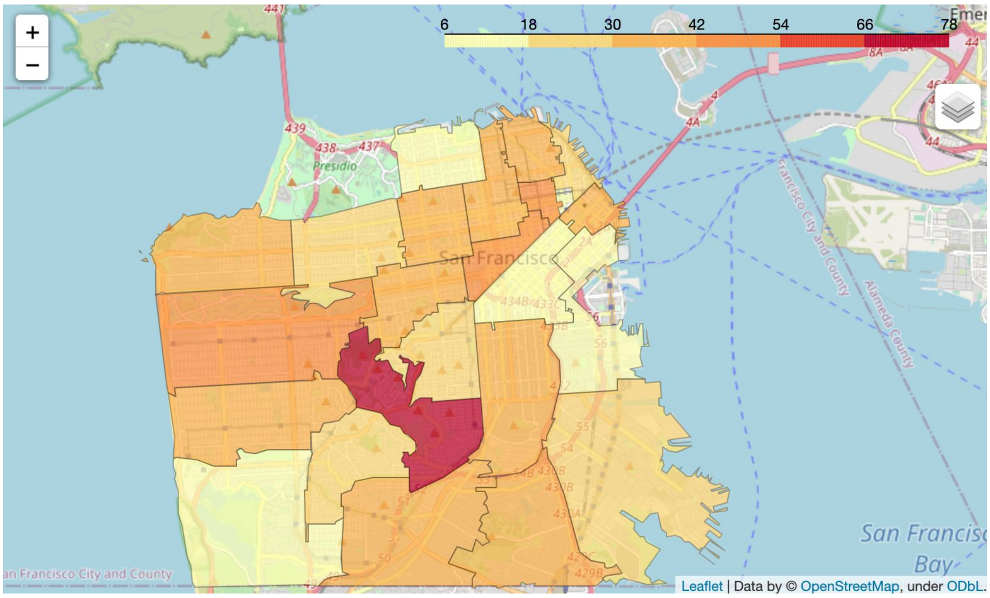


Figure A16: Percentage of businesses by zip code offering delivery services only

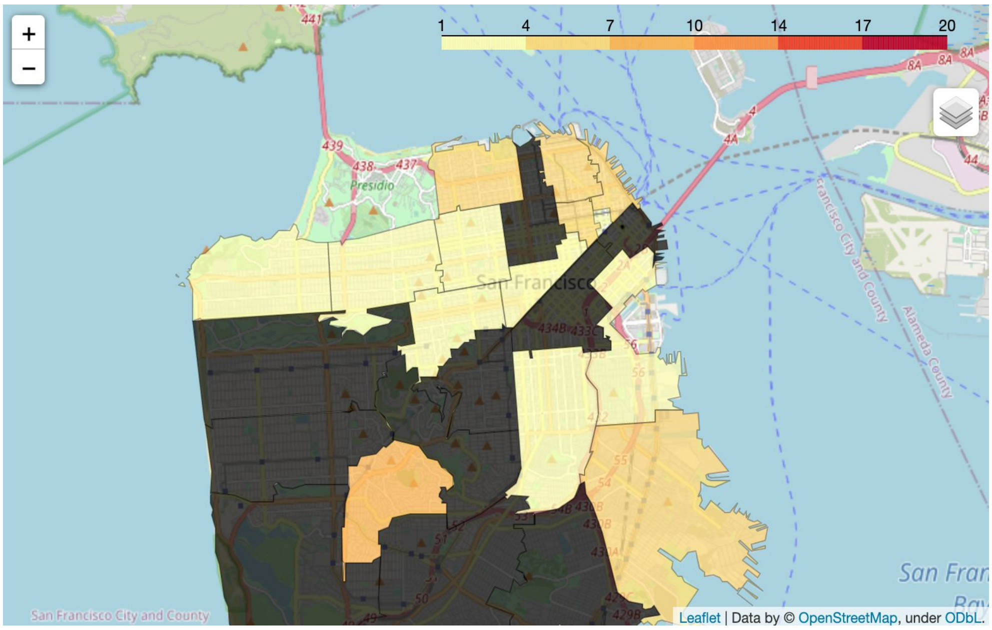


Figure A17: Percentage of businesses by zip code offering pick-up services only

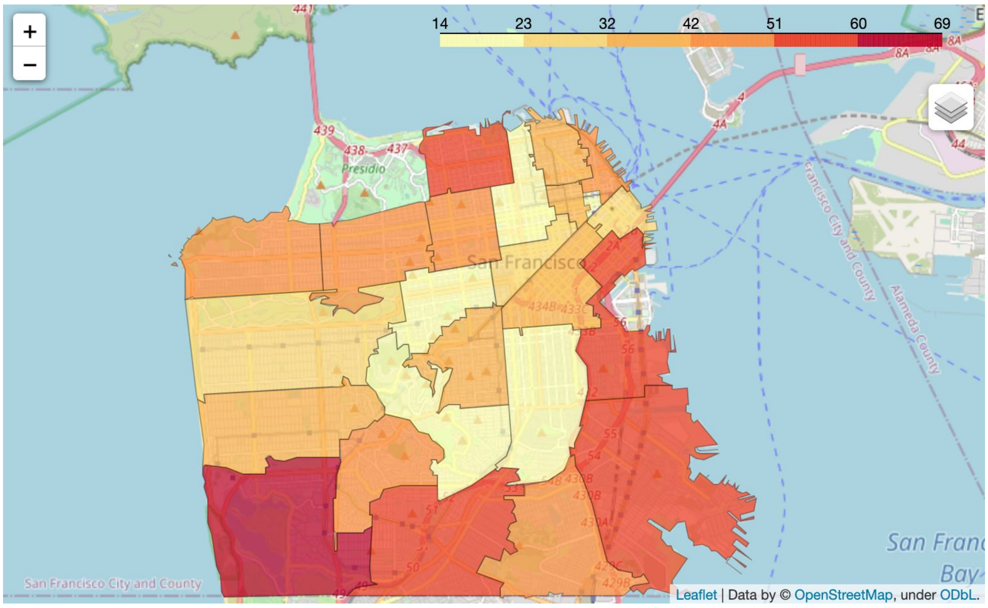


Figure A18: Percentage of businesses by zip code offering delivery and pick-up services

3.2: Miami-Dade County

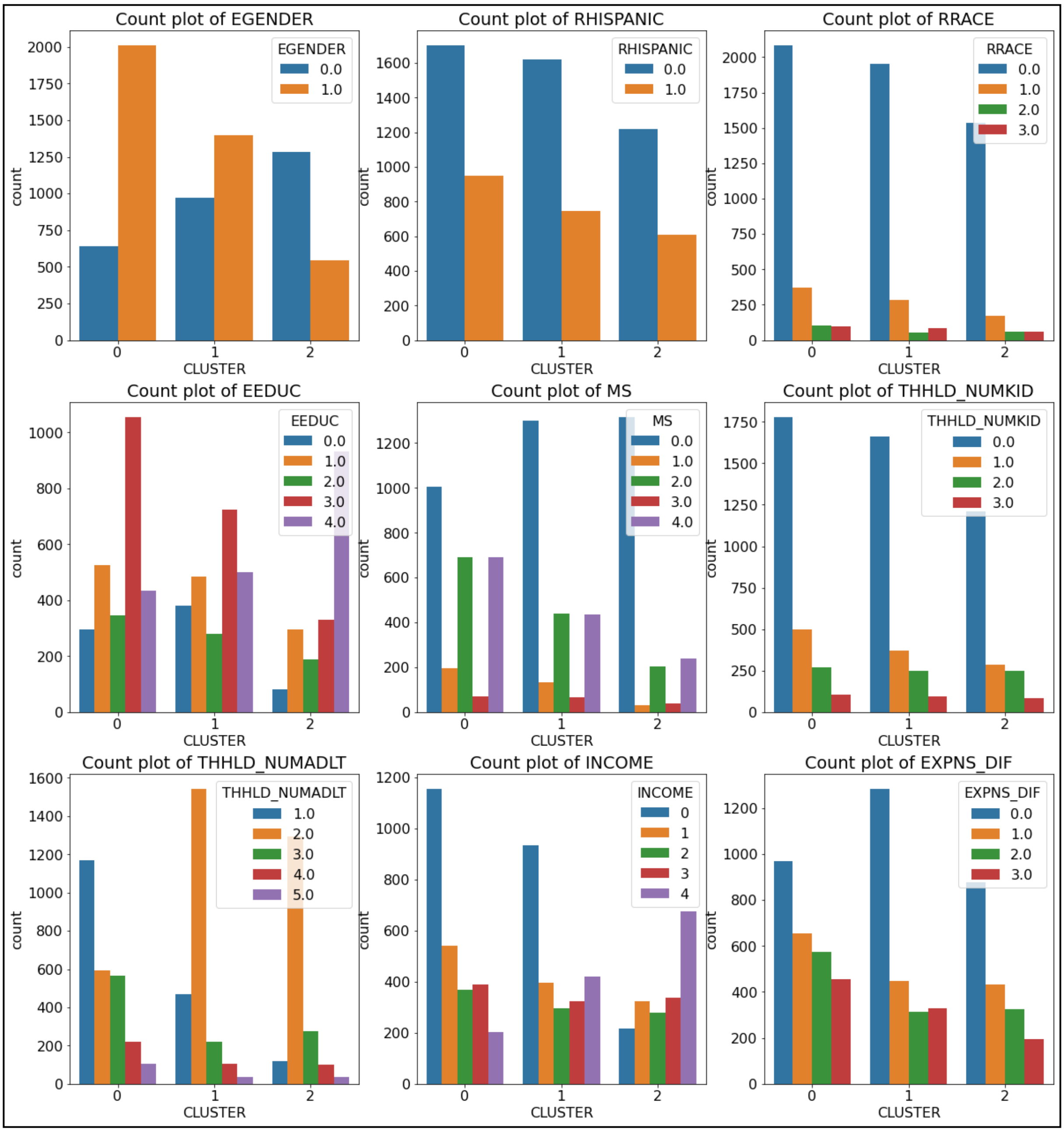


Figure A19: The distribution of demographics by cluster in Miami

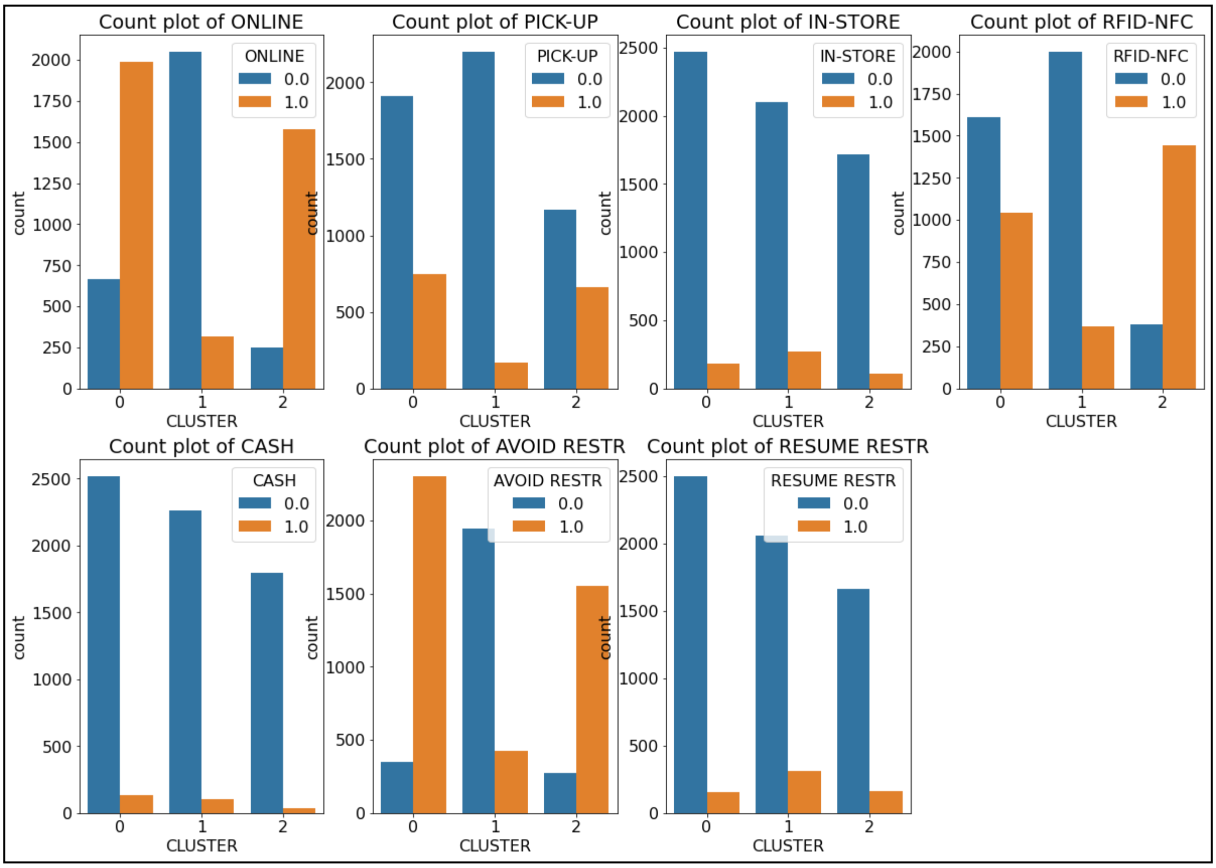


Figure A20: The distribution of shopping variables by cluster in Miami

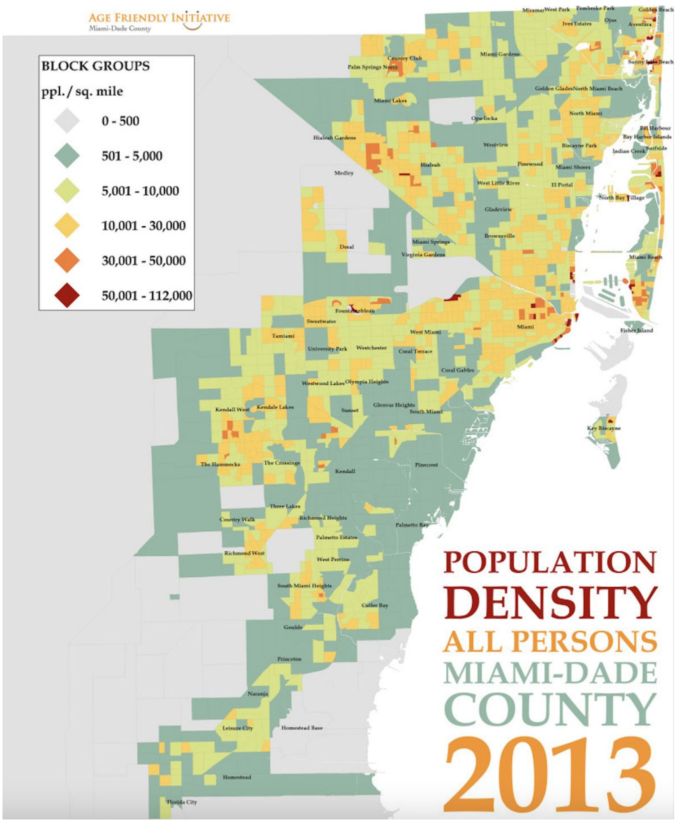


Figure A21: Referential population density in Miami county