

# How to help restaurants survive COVID-19

## Business Problem, Opportunity, and Impact

### Overview of industry, business, or problem

In 2020, 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 default 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, 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 dining, 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 dining (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.

## First glance at the Datasets

### Datasets

In this section, you should provide a description of the datasets used in your project (data types, sizes, sources, useful fields for your project. Highlight both the strengths and weaknesses of your datasets. Be aware of potential bottlenecks at the beginning is extremely relevant to make decisions early).

In this example, we use a table to organize the datasets and generate a list of features to be aware of (size, missing data, data types, variables connected to the project).

Datasets	Source	Description	Topics	Related topics
Household Pulse Survey	<a href="https://www.census.gov/data/experimental-data/products/household-pulse-survey.html">https://www.census.gov/data/experimental-data/products/household-pulse-survey.html</a>	<ul style="list-style-type: none"><li>• 17 weeks from April 23 to October 26, 2020.</li><li>• Surveyed people between 50k-100k per week. Number of variables: between 82 to 188, mostly categorical.</li><li>• Missing data designed as -88 and -99.</li><li>• Require the use of a data dictionary to translate the name of columns and categories.</li><li>• Demographics, spending, food, shopping, teleworking, trip trends variables.</li><li>• Dataset includes sub-variables</li></ul>	Shopping and purchase preferences.	Shopping modalities, payment modalities, resumed/avoided eating at restaurants. Use of credit cards, apps to buy online. Consumer preferences (prepared food vs ingredients to cook

		(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 specific analysis.		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)

## Methods and Models

### Visualizations

One key component of understanding purchase/payment preferences and likelihood to resume eating at restaurants during the pandemic is the identification of subgroups of population who share common predilections. In this example,

- Bar charts (univariate), stacked bar charts (multivariate) to explore distribution of the population by purchase/payment preference and demographics.
- Scatterplots and heatmaps to visualize correlations between shopping and demographic variables.
- Single and multi-line charts to track shopping preferences over time.
- Choropleth maps to visualize geographical pre-defined areas colored or patterned in proportion to a statistical aggregate variable. In this case, the maps will expose aggregate purchase variables by states, zip codes.
- Histograms to analyze the distribution of purchase variables and check if they distribute normally, data is skewed and presence of outliers.
- Boxplots to display the summary of behavioral variables (minimum, quartiles, median, maximum and presence of outliers).

### Which libraries to use

For initial phases of the project, pandas and pandas profiling allow an effective data exploration. For more detailed visualizations, scatterplots, boxplots, line charts and histograms are some of the [Matplotlib](#) functionalities. For interactive reports, [Seaborn](#), [Bokeh](#) and [Altair](#) could be good choices.

The exploratory data analysis is the first crucial step in modeling: provides summary level insights, reveals underlying patterns, missing data, outliers, biased, unbalanced data and relates the available data to the business opportunity and the modeling.

## Models

If you are going to explore a model, please provide a brief description why you think it is useful for your particular project. You must demonstrate you have a high-level understanding of the method. Don't use models that you don't understand or whose application is not even required to solve your project.

In this case, the differentiation of purchase profiles in the population is relevant for the project. Restaurant owners need to know community needs, purchase behaviors and who are their target customers. This project includes the use of:

- Statistical tests to find significance differences between subgroups.
- Feature engineering: Normalization, Scaling, One-Hot Encoding of variables before applying ML techniques.
- Unsupervised Machine Learning Models: Dimensionality Reduction, K-Means model to identify clusters or groups of customer profiles.

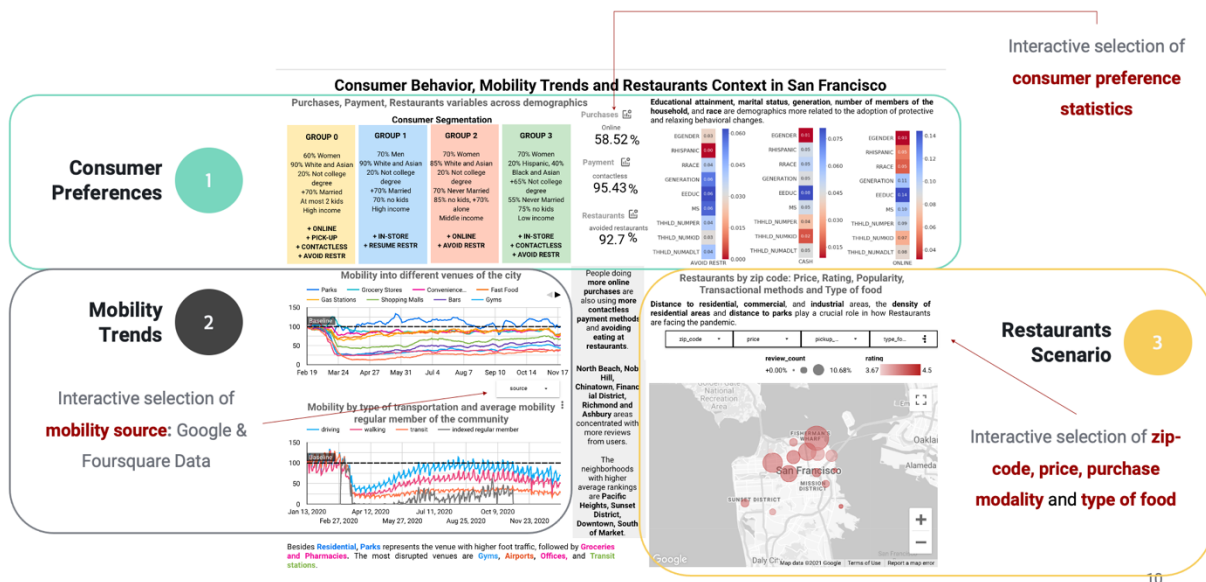
Other models/techniques to explore:

- Supervised ML for classification problems: Neural Networks, Decision Trees for classification, Logistic Regression, KNN, SVM, Naive Bayesian methods. Some of them explained [here](#).
- Supervised ML for regression problems: Neural Networks, Decision Trees for regression, Linear (Multi) Regression. Some of them explained [here](#).
- Unsupervised ML for classification problems: PCA, t-SNE, k-Means. Some of them explained [here](#).
- Feature Engineering: [Encoding techniques](#)
- Natural Language Processing techniques: Some techniques and applications explained [here](#).
- Deep Learning Models: Some applications in Python explained [here](#).

## Interface

In this section, you can include a picture of a hand-drawn sketch of the dashboard to show us that you have thought about what your final product will look like. Ideally your visualizations will have an interactive component.

In this example, the final front-end product will be an Analytics page for a specific location in the country with visualizations of the historical foot traffic data tracked during 2020, customers profiles based on unsupervised ML techniques, summary statistical purchase variables as key metrics, correlations between purchase and demographic variables and geographic visualizations with restaurants feature selectors. The end user can click on a particular mobility dataset to inspect foot traffic into different venues, display different purchase variables and restaurants features. Relevant insights from the analysis will be highlighted as text into the dashboard.

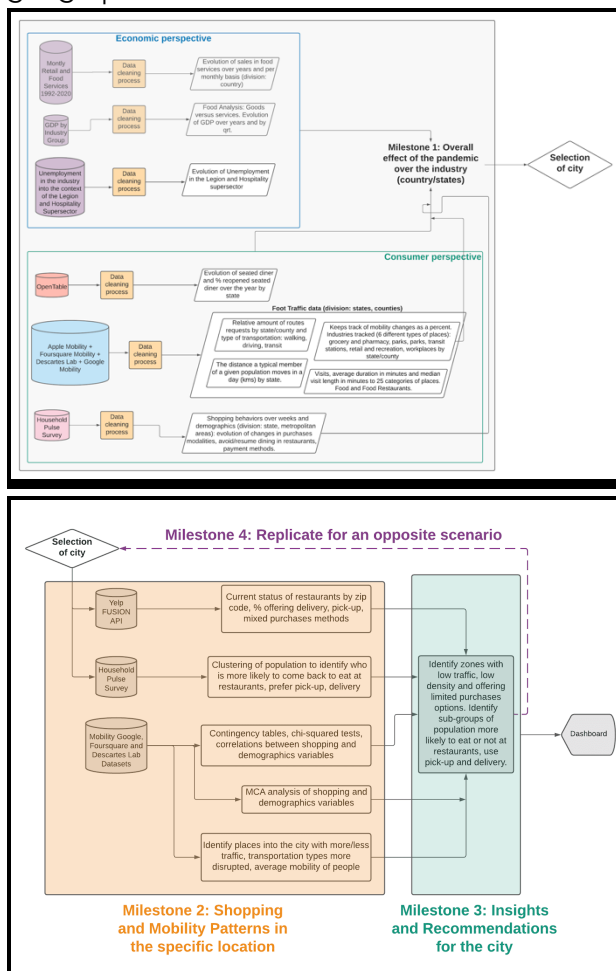


The above Figure is a finished interface. For early stages purposes, the scoping project doesn't require this deep level of detail. A hand-drawn or basic schema indicating the sections of your dashboard and what kind of visualizations and interactions are allowed to the user is enough. Planning and designing a dashboard (prototype) early is extremely important to have a destination in mind and avoid getting lost as you go. Starting with a destination gives you the ability to know where you are going with the project.

## Milestones

Your Version 1 must be achieved with 100% probability. Notice your Version 4 should be very difficult for you! We want to see you really challenge yourself in what the team could achieve. You should have minimum 3 versions and maximum 5 versions for your project. If you are a larger team with a more complicated project, it may be helpful to have more versions to stay organized.

In this example project, Milestone 1 (EDA) is the overall effect of the pandemic over the restaurant industry during 2020 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 (changes in consumer behaviors, mobility patterns related to the acquisition of food). This macro analysis pushes the project to Milestone 2 (in-depth EDA), where a specific geographic region is chosen to find out the status of restaurants, customer behaviors and mobility patterns within specific places in the city and counties. Milestone 3 (statistics and modeling) consist of statistical tests and application of unsupervised Machine Learning to get insights and recommendations for local businesses. Milestone 4 is a geographical extension to another location on opposite spectrum.

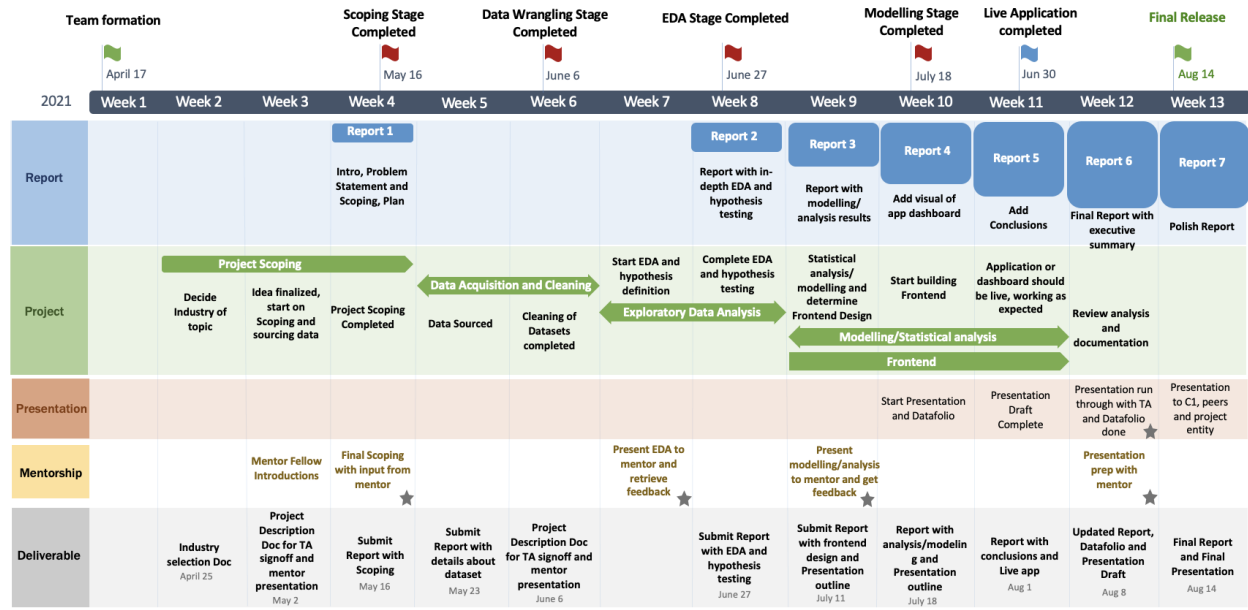


Example of Milestone versions

## Timeline

Ideally tell us who will do what, especially if you're part of a larger team (5+ people). If you use the following table as reference, you can add additional columns to specify which Milestone you are working on every week and detailed tasks for different team members.

## DS4A Project Timeline of Deliverables



You can zoom in on the above timeline by checking out the [original slide](#).

Date	Deliverable	Details
Week 1	Team Formation	
Week 2	Work on idea formation	
Week 3	Idea should be finalized & Start on Scoping	
Week 4	Project Scoping Completed	
Week 5	Datasets sourced	
Week 6	Basic EDA/Cleaning of datasets completed	
Week 7	In-depth EDA, jupyter analysis, mockup of frontend	
Week 8	Frontend Design	
Week 9	Front End Complete	

Week 10	Application infrastructure complete	
Week 11	Project and Datafolio complete	
Week 12	Presentation rehearsed and ready	

## Concerns

In this case example, concerns are related to:

- How waves of new COVID cases will likely affect the next series of data. Mobility and purchase preference data was updated monthly in the data sources and the team was aware that they needed to collect the updated data for tracking variables over time.
- Some datasets missed key elements to merge information appropriately. Doing an effective first glance of the datasets, the team was able to drop datasets with meaningful missing information in the early stages and only focus on available and updated data.
- Work distribution: emphasis on individual strengths and weaknesses to determine how to support the rest of the team. Communication, commitment and specific deadlines for mini tasks assigned to each team member.