

A decorative graphic on the left side of the slide, consisting of a network of white lines and small circles on a blue gradient background, resembling a circuit board or data flow diagram.

A NEW BUSINESS IN SAN FRANCISCO

IBM DATA SCIENCE CAPSTONE PROJECT

BUSINESS PROBLEM

- San Francisco is one of the most expensive cities to live in America, yet it also offers qualities that would appeal greatly to businesses:
 - High median income = more disposable income for potential customers to spend
 - Compact city layout = greater density of customers
 - Proximity to Silicon Valley = large population of tech savvy customers
- With this in mind, we wish to look into which area of San Francisco is most ideal for a new service-based industry to locate.

RESEARCH FOCUS AND JUSTIFICATION

- Instead of analyzing by neighborhoods, we analyzed by zip codes
 - San Francisco's neighborhoods are too numerous and have significantly overlapping areas, making analysis difficult. We use zip codes instead because they are geographically distinct regions.
- We analyze the number and type of venues in each zip code location
 - By looking at the number of and type of venues in each location, we are able to get a sense of the density of venues in those locations. Higher density of venues would suggest a greater density of customers for a new business.

METHODOLOGY

- Input data from zipdatamaps.com, which includes coordinates for each zip code in San Francisco
- Using Foursquare API to identify the venues in each zip code and which types are most common.
- One hot incoding and Kmeans clustering is used to identify most common types of venues in each zip code.
- Folium mapping is used to visualize the data.

CONCLUSION

- The data seems to indicate that the highest density of venues is located in the northeastern region of the city (specifically zip codes 94102, 94103, 94108, and 94109). This corresponds to the Financial and Downtown Districts.
- This suggests that these two areas are most suitable for a prospective business in San Francisco. That being said, while not considered for this study, real estate prices in these regions is very high. It therefore may be more prudent to consider locating a new business on the outer edges of these zip codes instead.