

## Open Capstone Project Proposal – Market research

Yeh, Margaret

### Description:

In business, the startup cost is a significant part of the investment. It is important to figure out the best locations early to invest in to result in a profit in the future. This project uses data publicly hosted from the US Census Bureau to determine locations that are more likely to show strong economic growth in the future. Metrics I plan to examine include: Increasing salary, increasing population, location, education, healthcare options, restaurant and supermarket statistics

As an additional benefit, this project may be able to determine areas which show trends of economic recession. This can be useful as an indicator of places to avoid or to move away from. It can also be an indicator for places that will need additional help in the future.

The dataset for this project will be stored onto Azure Blob, which can be scaled upwards in the future if necessary.

### Criteria:

The data collected will need to span a significant timeframe to be able to predict future trends. I will try to aim the total data amount to around 15 GB in size.

The data ingestion will need to be rerun on a yearly basis to include incoming data for the new year.

### Plan:

Use the US Census Bureau API to get a JSON with specified categories for each year

Create Spark Dataframes with a specified schema, load the data into it, clean it, and save the result as parquets in Azure blob under folders of each specified year

Load parquets from Azure blob and run queries on it to find key locations

Use Tableau Public to generate a dashboard to show results