

CMPT 732 Project Proposal

1. Abstract

It can be challenging for business owners to decide on the best location for their stores, the products they will sell, and how to increase sales when starting a new or expanding an existing business in a new city.

Therefore, by examining the open data set assembled by the government, our initiative intends to inform them with knowledge that can assist them in making decisions.

We plan to test this strategy on the Liquor sales dataset presented by Iowa Govt Website. After obtaining the data, we will need to clean it, extract-transform-load it into a Data Warehouse and then eventually into a Data Mart, and then generate reports like grouping the most popular drink by region. And the final stage will be to visualize the data and make it presentable to the clients.

2. Datasets

Iowa Liquor Sales:

<https://gist.github.com/dannguyen/18ed71d3451d147af414>

<https://data.iowa.gov/Sales-Distribution/Iowa-Liquor-Sales/m3tr-qhgy>

Iowa Liquor Stores:

<https://data.iowa.gov/Regulation/Iowa-Liquor-Stores/ykb6-ywnd>

Iowa Liquor Product Portfolio:

<https://data.iowa.gov/stories/s/nctk-7iu4>

COVID-19 Hospitalization Reporting Data:

https://gis.cdc.gov/grasp/COVIDNet/COVID19_3.html

3. Technologies

We plan to use Hadoop, Spark, HDFS, Cassandra, Celery, Angular/React.