Project 2 proposal

**Contributors**

Maurice Jenkins

Collin Seebeck

Clifford Ouma.

Megan Williams

For project 2, the team has decided to collect data on shark attacks globally and

then only filtered for USA states, we are also going to get ice cream data in the US

states.

Data Sources.

•Shark attacks:

https://www.sharkattackfile.net/incidentlog.htm,this had the

option to save as CSV right from the website

•Google data for ice cream stores near attack sites: API call to googleapis.com/maps

We will be reading the shark attack data using pandas and then cleaning the data

by only selecting the columns we need, dropping NAN values and dropping the

index. This data will then be imported into PostgreSQL for manipulation.

We are looking to join the ice cream location and shark attack based on the locations and making a comparison there

\*\*This is amended from our original proposal based on pay rates for needed weather data and the

**Final Technical Report**

**Extraction**

We pulled data on international shark attacks reported to a central reporting repository via [www.sharkattackfile.net/incidentlog.htm](http://www.sharkattackfile.net/incidentlog.htm). The data available was in CSV format and extensive.

We then used a Google API to locate ice cream shops in within roughly 1 mile of each location. This was done using a Google API call.

**Transformation process**

Shark Attack Data

The extensive data regarding shark attacks is of interest, however, since there appears to be no required format for submission and goes back over 100 years, extensive clean-up was necessary. See attached Source to Target document

Ice Cream Shop Data

We used the latitude and longitudes of each location in the shark attack csv to locate ice cream shops within roughly a one-mile radius to then create a new data frame to be loaded into our database. Little transformation was done with this

**Loading process**

Both data sets were cleaned using pandas.

The data from the shark attacks was already in a CSV format. The cleaned version was loaded into postgresql after creating the table shark\_attacks

The data from Google API was also saved into a CSV file after the transformation process using jupyter lab. This was then also loaded into postgresql after creating the table ice\_cream.

Once in postgres we joined those two tables into ice\_cream\_sharks on the location of attack.