# Big City Vacation - Deliverable 1

## Team:

Laura Clark (circle), Christine Mitchell (square), and Josh Stephens (triangle)

## Story:

Before you grab your walking shoes, cameras, and water bottle, to explore “The Big Apple” or “The City”, you will need to plan for your big city vacation adventure and we are here to help.

The first question you may ask yourself is, "when should I book my trip to NYC or London?" Our team can help answer that question by accounting for weather. We will create a model to predict the average maximum temperature for each month. Based on your preferences, we can then provide suggestions as to when to schedule your vacation.

But that is not all, we would like to provide additional resources to get the most out of your vacation by creating a layered map that can be used to build out your daily itenary.

 First, we would like to provide optional hotel accommodations that meet a minimum rating threshold.

 Second, we would like to provide suggested points of interests and group them together by area so you can make the most out of each day.

 Third, we realize how important it is to find a good restaurant near your activities and lodging as well as how it can add to the cultural experience, so we would like to display “What restaurants are in the area and how good they are” on the layered map.

So, let’s get started.

## Tools & Technologies:

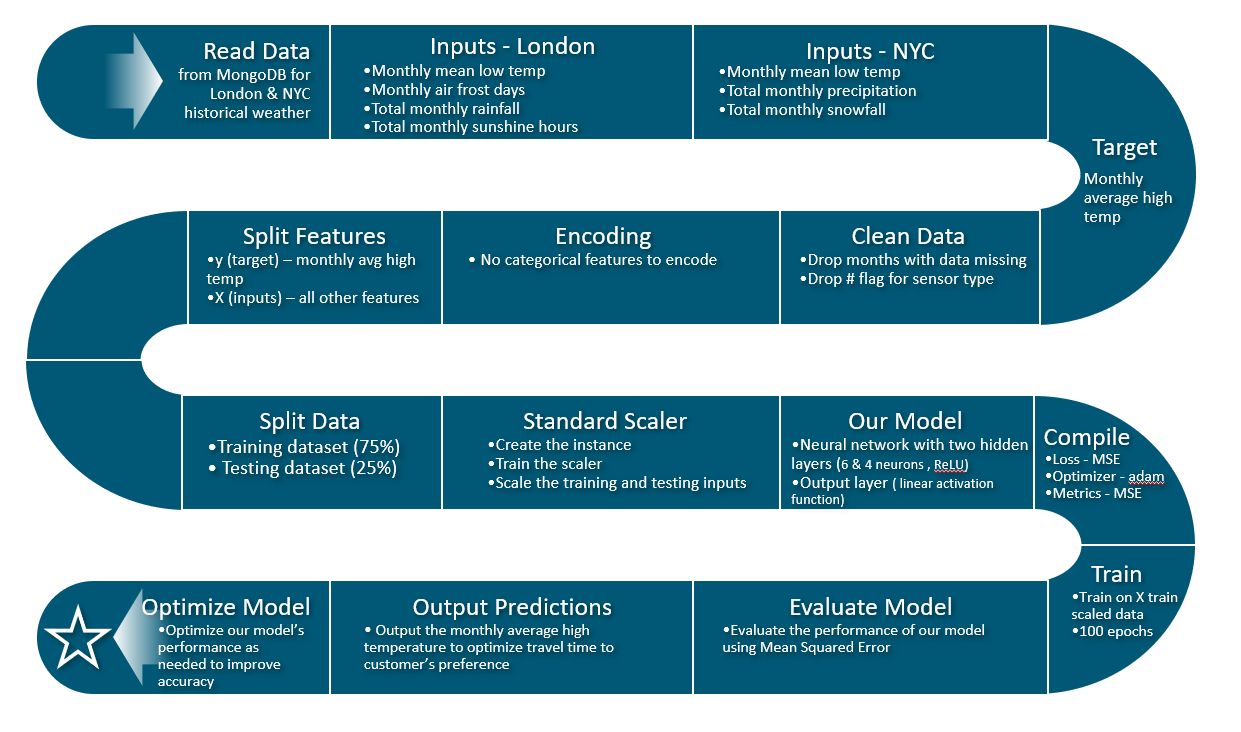
 Excel for initial csv file data exploration and some cleaning.

 Jupiter Notebook using PANDAS & Python to further explore data, clean the data, and train the model.

 Data Storage: Mongo DB utilizing postregresql to store our weather data for predictive modeling.

 Tableau to develop our interactive dashboard and create a layered map with points of interest for a vacation in NYC or London. Points of interest will include hotels, eateries, and attractions and will be loaded with csv files.

 Machine Learning Neural Network Model trained to predict the average maximum temperature for each month based on the scaled inputs like sunlight hours, air frost days, minimum temperature, precipitation, and snowfall (independent variables). This model was selected in an attempt to leverage the flexibility of the various hidden layers, neurons, activation functions, etc. The model is evaluated based upon loss and mean squared error (MSE). We’re hopeful that our model can be used to present months that match a travelers desired high temperature.

[](https://user-images.githubusercontent.com/82730954/132999771-d43477d1-127c-437c-9d9d-f4c265174024.PNG)

## Data Sources:

### Weather Data:

LONDON:  
Source: <https://www.metoffice.gov.uk/pub/data/weather/uk/climate/stationdata/heathrowdata.txt>  
More information: <https://www.metoffice.gov.uk/research/climate/maps-and-data/historic-station-data>  
  
The dataset contains monthly weather data for Heathrow (London Airport) from January 1948 - August 2021. This dataset contains 884 rows and 7 columns of labels and values. Each month includes details such as:

 Mean daily maximum temperature

 Mean daily minimum temperature

 Total rainfall

 Air frost days

 Total monthly sunshine (the device for measuring sunshine changed 9/2005. Removed # indicator from sunshine data in Excel using substitute formula suggested here: <https://spreadsheetplanet.com/remove-specific-character-from-string-excel/>)  
The data was loaded into a DataFrame and grouped by month. Averages and minimum and maximum values were generated.

NEW YORK CITY:  
Source: <https://www.weather.gov/wrh/Climate?wfo=okx>  
Monthly summarized data was captured for NY-Central Park Area from the site above for 1948 - 2021. This dataset has 884 rows and 6 columns of labels and values. Each month includes the following details:

 Mean daily maximum temperature

 Mean daily minimum temperature

 Total precipitation

 Total snowfall  
Using vlookups, the data was consolidated from multiple tables and worksheets into one CSV file that mimicked the format of the London data. The dataset was loaded into a DataFrame and grouped by month. Averages and minimum and maximum values were captured.

### Hotels:

The property name, address, and rating will be utilized in our layered map pop up marker. The coordinates are necessary to include them on our map.

LONDON:  
<https://data.world/promptcloud/customer-of-reviews-of-london-based-hotels>  
Original data set included the property name, review rating, review title, review text, reviewer location. Latitude and longitude coordinates were added to the dataset using google maps at the source.

NEW YORK CITY:  
<https://www.kaggle.com/gdberrio/new-york-hotels>  
Data set includes the property name, address, latitude, longitude, star\_rating, high rate, and low rate.

### Points of Interest:

LONDON:  
It took a while to find a website with the desired data, and that was <http://tour-pedia.org/about/datasets.html>  
Although it needed heavy editing and some geocordinates from an API, this had the information about the name and address.

NEW YORK CITY:  
Data.gov was used and these were all the datasets pulled

 <https://catalog.data.gov/dataset/new-york-city-museums>

 <https://catalog.data.gov/dataset/new-york-city-art-galleries>

 <https://catalog.data.gov/dataset/new-york-city-water-trail-kayak-and-canoe-launch-sites>

 <https://catalog.data.gov/dataset/directory-of-temporary-public-art>

 <https://catalog.data.gov/dataset/nyc-parks-monuments>  
Although the format differed between most of the datasets, they all had a name, most had an address, and all had latitude and longitude in one form or another

### Restaurant Reviews:

<https://www.kaggle.com/yelp-dataset/yelp-dataset>

## Logistics:

### Communication Protocal:

 Slack Channel – Group 2

 Exchange personal numbers for group texts in the event of emergency or no response after 24 hours in slack.

### Manage Project:

 Git Hub Repository: <https://github.com/ChristineMitchell/Final_Project_Travel>

 Git Hub Task List created via Issues tab of the repository

### Meeting:

 Class time

 As needed, Josh Stephens host zoom meetings.