# **HW: Surfs Up! (required) with my Notes:**

****

Congratulations! You've decided to treat yourself to a long holiday vacation in Honolulu, Hawaii! To help with your trip planning, you decided to do some climate analysis on the area. Because you are such an awesome person, you have decided to share your ninja analytical skills with the community by providing a climate analysis api. The following outlines what you need to do.

## **Step 0 - Project**

Decide who will be in your group for your final project (you can be by yourself if you’d like) and what your project will be.

In addition, identify the data sources for your project.

Fill out this form: <https://docs.google.com/forms/d/e/1FAIpQLSepgTX5XzOP3zuM1d_WgkLKkmlnA2jBNx9zqnE8wOR8CUhYJw/viewform?usp=sf_link>

Check out the previous projects here: <https://docs.google.com/spreadsheets/d/1Sg7I8NlhzIEi8WLYo1_xRoA5Vu-lePeESqYER5tgg60/edit#gid=2073386660>

## **Step 1 - Data Engineering**

The climate data for Hawaii is provided through two CSV files. Start by using Python and Pandas to inspect the content of these files and clean the data.

The homework instructions with updates are at:

[https://docs.google.com/document/d/1c8w5U6lZGyzBga7q50YC3nROAvDz8blxIOOepMTiUxo/edit#](https://docs.google.com/document/d/1c8w5U6lZGyzBga7q50YC3nROAvDz8blxIOOepMTiUxo/edit)

Here are the files: <https://drive.google.com/open?id=0Bz0Wzew04n0uTTJzVVB3ajV1OUU>

* Create a Jupyter Notebook file called data\_engineering.ipynb and use this to complete all of your Data Engineering tasks.
* Use Pandas to read in the measurement and station CSV files as DataFrames.
* Inspect the data for NaNs and missing values. You must decide what to do with this data.
* Save your cleaned CSV files with the prefix clean\_.

***My Notes on Missing Data:***

*Stations with 2010 - 2017 data*

*KANEOHE 838.1, HI US*

*KUALOA RANCH HEADQUARTERS 886.9, HI US*

*MANOA LYON ARBO 785.2, HI US*

*PEARL CITY, HI US*

*WAIHEE 837.5, HI US*

*WAIKIKI 717.2, HI US*

*WAIMANALO EXPERIMENTAL FARM, HI US*

*Stations with 2010 - 2015 data*

*HONOLULU OBSERVATORY 702.2, HI US*

*UPPER WAHIAWA 874.3, HI US*

*8 out of 9 stations have blank prcp values for 1447 records with blanks.*

*USC00511918 47 blank prcp*

*USC00513117 13 blank prcp*

*USC00514830 265 blank prcp*

*USC00516128 128 blank prcp*

*USC00517948 689 blank prcp*

*USC00518838 169 blank prcp*

*USC00519397 39 blank prcp*

*USC00519523 97 blank prcp*

*The missing data is spread out over the years of data so removing them doesn't seem to be a large concern as opposed to if the missing data was concentrated in a station or year.*

## 

## **Step 2 - Database Engineering**

Use SQLAlchemy to model your table schemas and create a sqlite database for your tables. You will need one table per weather station. Use the station name as the table name.

* Create a Jupyter Notebook called database\_engineering.ipynb and use this to complete all of your Database Engineering work.
* Use Pandas to read your cleaned measurements and stations CSV data.
* Use the engine and connection string to create a database called hawaii.sqlite.
* Use declarative\_base and create ORM classes for each table.
  + You will need a class for Measurement and for Station.
  + Make sure to define your primary keys.
* Once you have your ORM classes defined, create the tables in the database using create\_all.

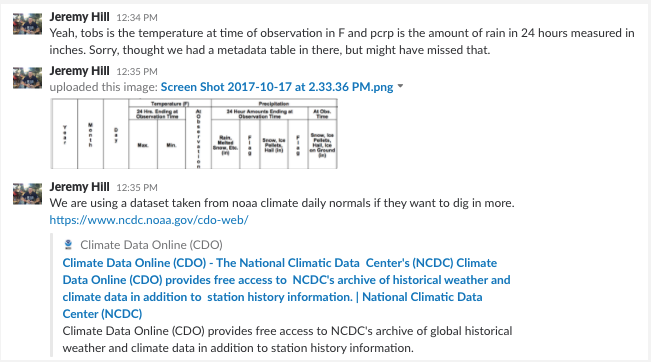
***My Notes:***

*Class activities on sql Alchemy and ORM started on 09/26.*

## **Step 3 - Climate Analysis and Exploration**

You are now ready to use Python and SQLAlchemy to do basic climate analysis and data exploration on your new weather station tables. All of the following analysis should be completed using SQLAlchemy ORM queries, Pandas, and Matplotlib.

* Create a Jupyter Notebook file called climate\_analysis.ipynb and use it to complete your climate analysis and data exporation.
* Choose a start date and end date for your trip. Make sure that your vacation range is approximately 3-15 days total.
* Use SQLAlchemy create\_engine to connect to your sqlite database.
* Use SQLAlchemy automap\_base() to reflect your tables into classes and save a reference to those classes called Station and Measurement.



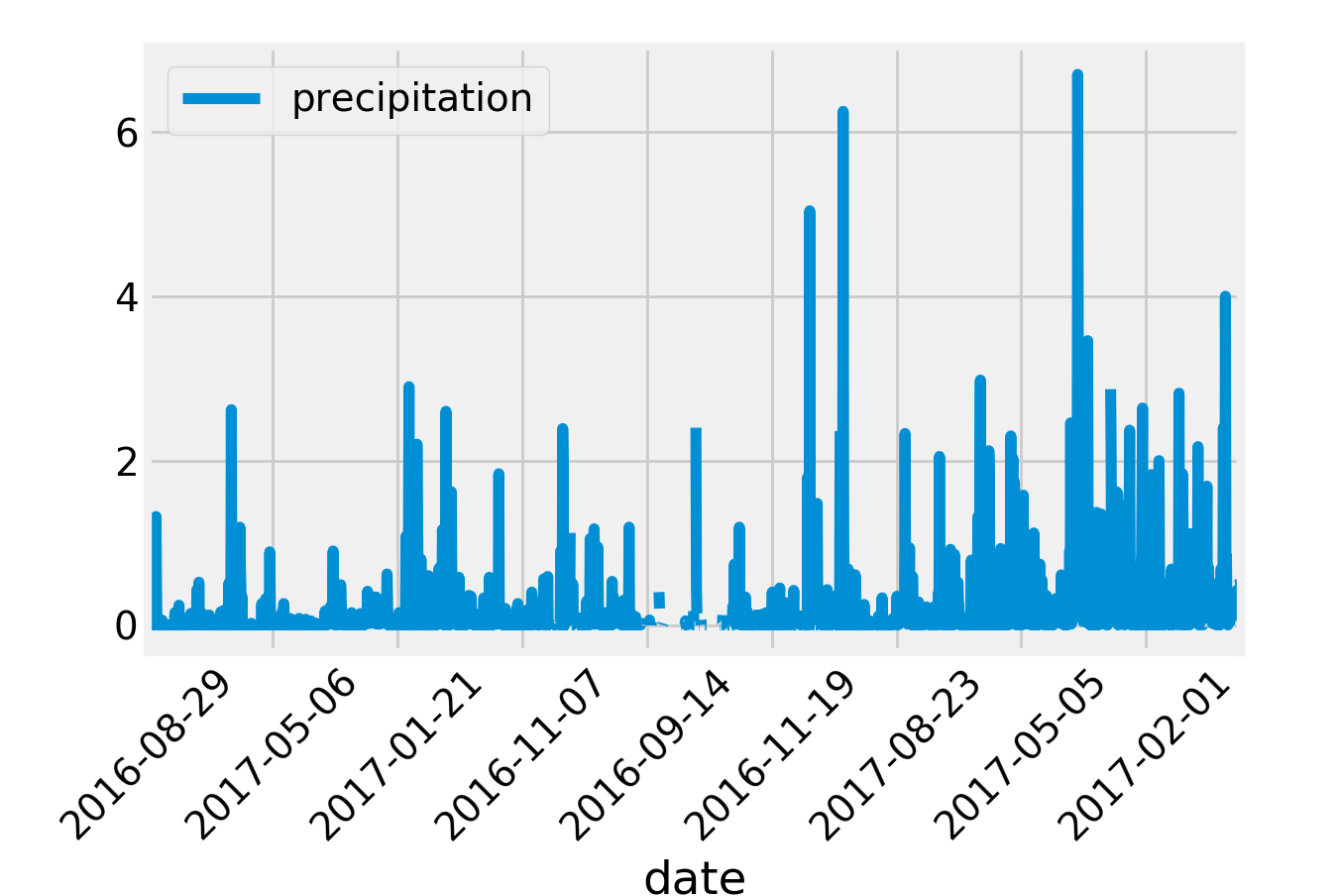
**df.plot:**

<https://pandas.pydata.org/pandas-docs/stable/generated/pandas.DataFrame.plot.html>

<https://pandas.pydata.org/pandas-docs/stable/visualization.html>

### **Precipitation Analysis**

* Design a query to retrieve the last 12 months of precipitation data.
* Select only the date and prcp values.
* Load the query results into a Pandas DataFrame and set the index to the date column.
* Plot the results using the DataFrame plot method.

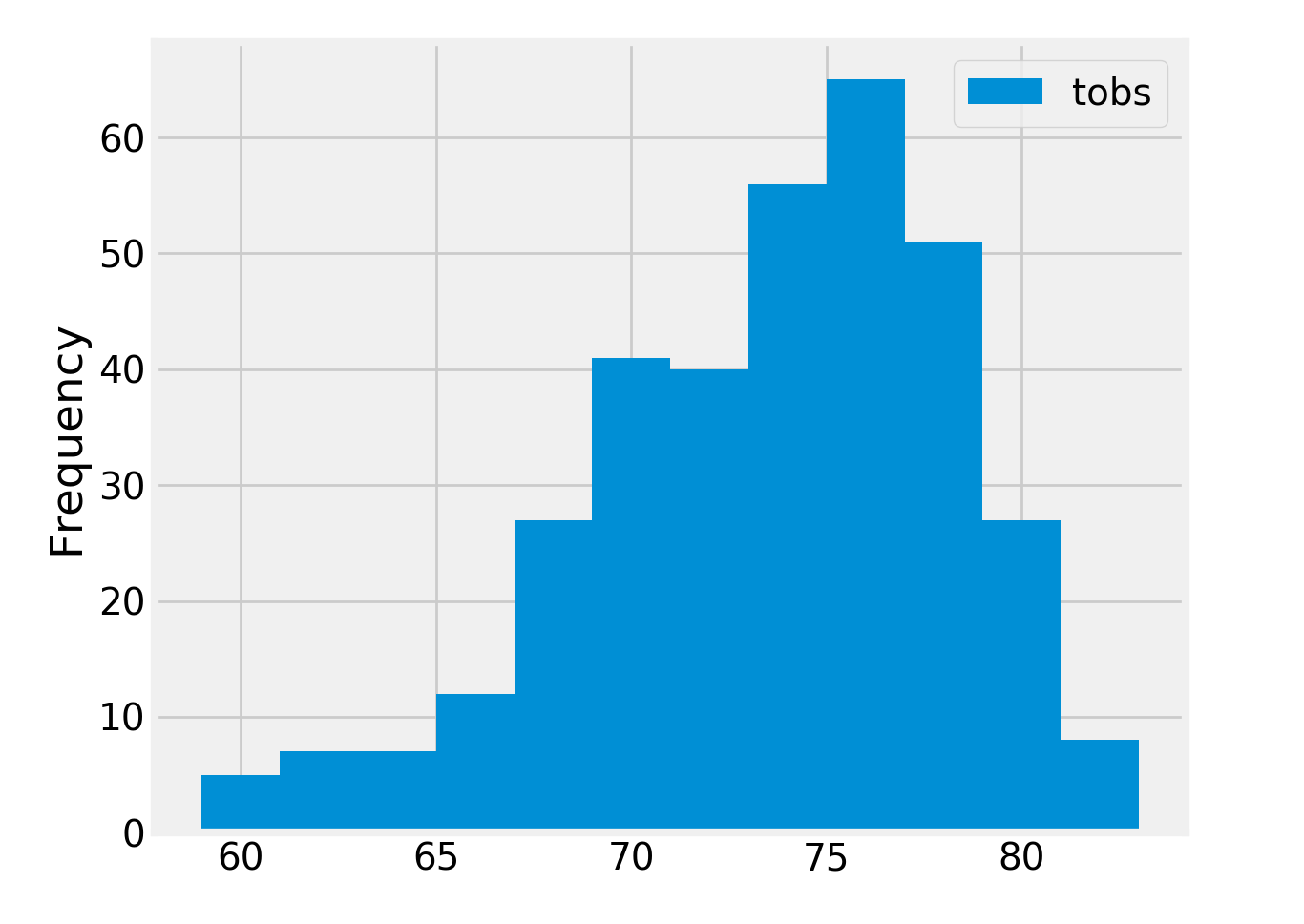


* Use Pandas to print the summary statistics for the precipitation data*.* 
  + *From the dataframe, run min, max, mean, sum, and describe().*

<https://chrisalbon.com/python/pandas_dataframe_descriptive_stats.html>

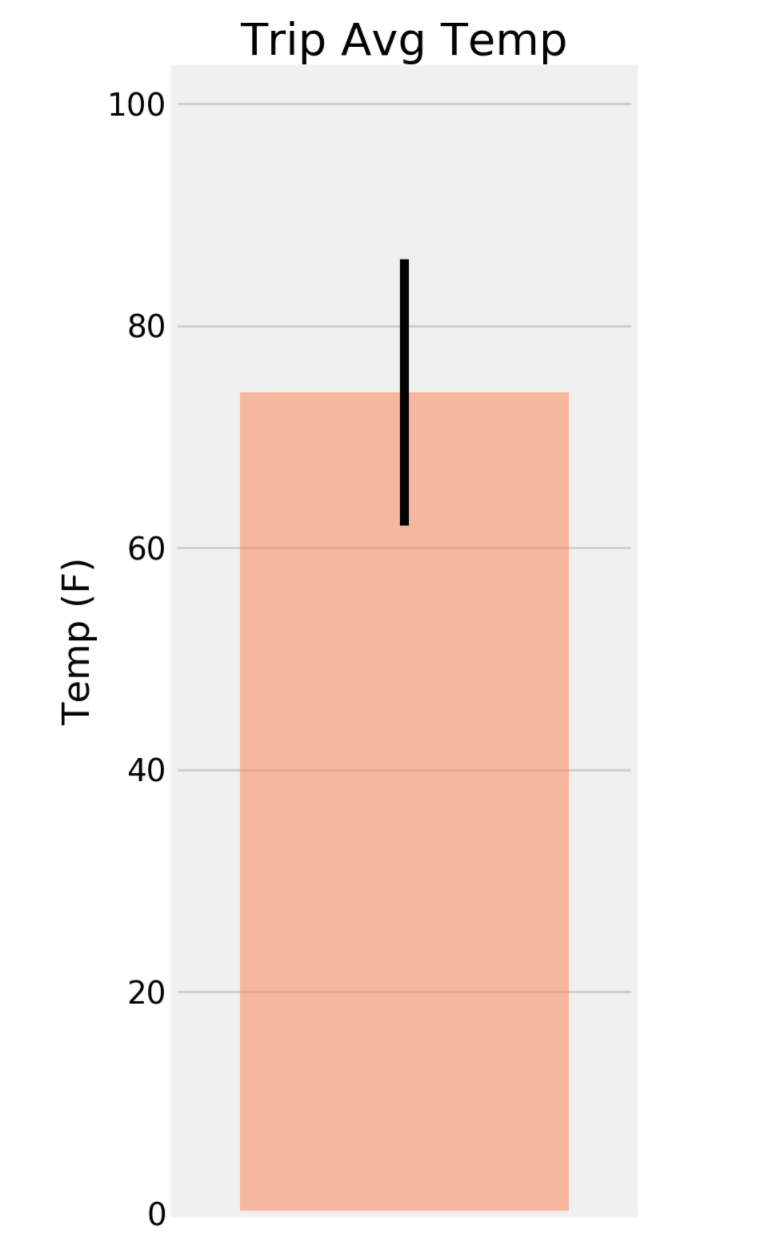
### **Station Analysis**

* Design a query to calculate the total number of stations.
* Design a query to find the most active stations.
  + List the stations and observation counts in descending order
  + Which station has the highest number of observations?
* Design a query to retrieve the last 12 months of temperature observation data (tobs).
  + Filter by the station with the highest number of observations.
  + Plot the results as a histogram with bins=12.



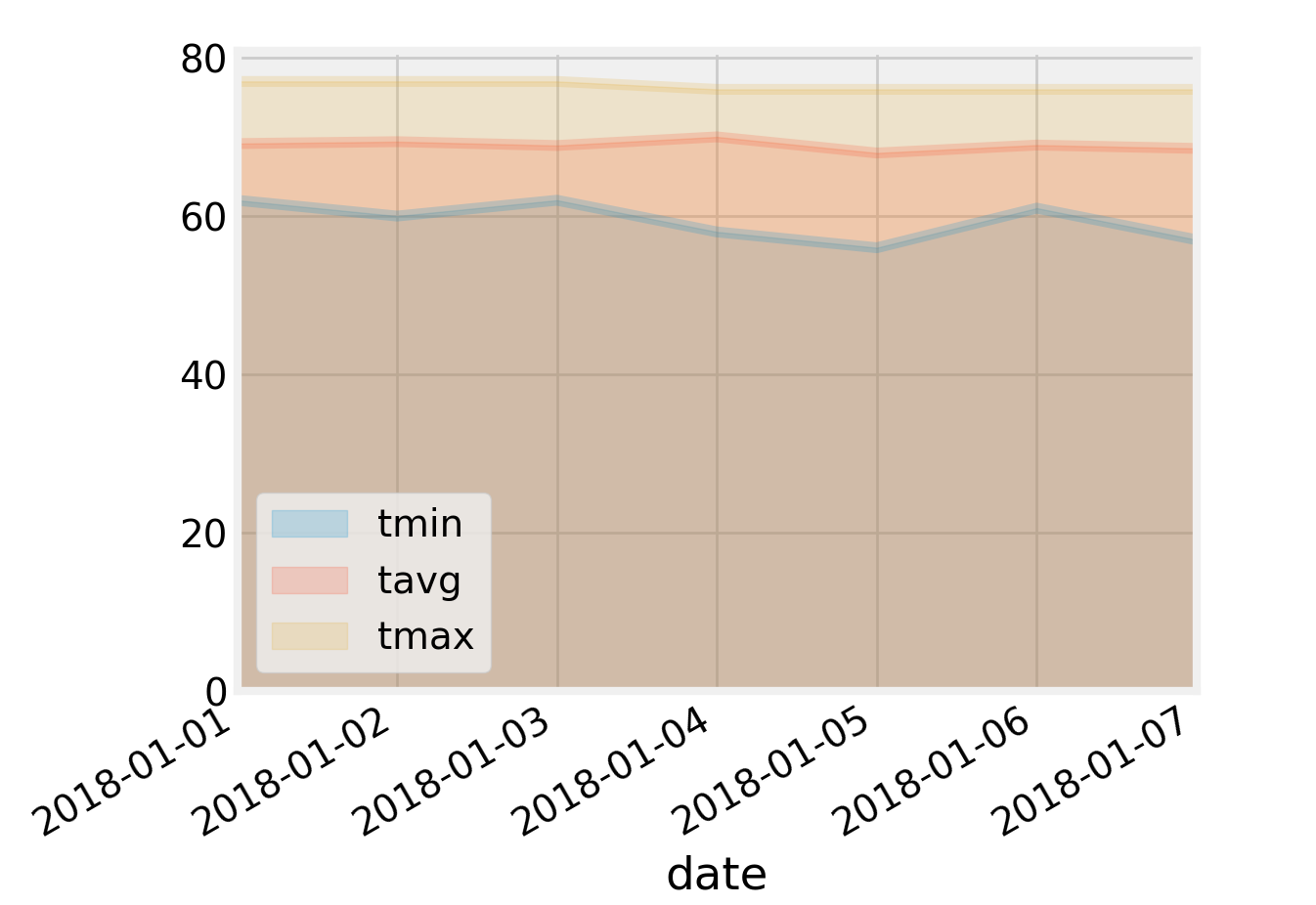
### **Temperature Analysis**

* Write a function called calc\_temps that will accept a start date and end date in the format %Y-%m-%d and return the minimum, average, and maximum temperatures for that range of dates.
* Use the calc\_temps function to calculate the min, avg, and max temperatures for your trip using the matching dates from the previous year (i.e. use "2017-01-01" if your trip start date was "2018-01-01")
* Plot the min, avg, and max temperature from your previous query as a bar chart.
  + Use the average temperature as the bar height.
  + Use the peak-to-peak (tmax-tmin) value as the y error bar (yerr).



### **Optional Recommended Analysis**

* The following are optional challenge queries. These are highly recommended to attempt, but not required for the homework.
  + Calculate the rainfall per weather station using the previous year's matching dates.
* Calculate the daily normals. Normals are the averages for min, avg, and max temperatures.
  + Create a function called daily\_normals that will calculate the daily normals for a specific date. This date string will be in the format %m-%d. Be sure to use all historic tobs that match that date string.
  + Create a list of dates for your trip in the format %m-%d. Use the daily\_normals function to calculate the normals for each date string and append the results to a list.
  + Load the list of daily normals into a Pandas DataFrame and set the index equal to the date.
  + Use Pandas to plot an area plot (stacked=False) for the daily normals.



## **Step 4 - Climate App**

Now that you have completed your initial analysis, design a Flask api based on the queries that you have just developed.

* Use FLASK to create your routes.

### **Routes**

* /api/v1.0/precipitation
  + Query for the dates and temperature observations from the last year.
  + Convert the query results to a Dictionary using date as the key and tobs as the value.
  + Return the json representation of your dictionary.
* /api/v1.0/stations
  + Return a json list of stations from the dataset.
* /api/v1.0/tobs
  + Return a json list of Temperature Observations (tobs) for the previous year
* /api/v1.0/<start> and /api/v1.0/<start>/<end>
  + Return a json list of the minimum temperature, the average temperature, and the max temperature for a given start or start-end range.
  + When given the start only, calculate TMIN, TAVG, and TMAX for all dates greater than and equal to the start date.
  + When given the start and the end date, calculate the TMIN, TAVG, and TMAX for dates between the start and end date inclusive.

## **Hints**

* You will need to join the station and measurement tables for some of the analysis queries.
* Use Flask jsonify to convert your api data into a valid json response object.

## **Copyright**

Coding Boot Camp © 2017. All Rights Reserved.