**CHALLENGE 10**

**REPORT**

**Introduction**:

This study is regarding weather data from Honolulu County in Hawaii. The data is analyzed in preparation for a vacation trip to Honolulu. Data is obtained from the database ‘hawaii.sqlite’. SQLAlchemy and Python are used to perform ORM queries. Using the query results, an API with various query options is created.

**Data**:

**Part 1**

Pandas and Matplotlib are used to analyze and plot graphs.

Nine (9) weather stations are in the database. The following graph shows the precipitation in inches recorded by all stations over the period of the last 12 months in the database.

A graph of blue lines

Description automatically generated

The following table and graph are the temperature statistics for data collected by the most active weather station,

USC00519281, from August 18, 2016, to August 18, 2017.

|  |  |
| --- | --- |
| Temperature Statistics for Station **USC00519281** | |
| mean | 73.20 °F |
| median | 74.00 °F |
| standard deviation | 4.76 °F |
| mode | 74.00 °F |
| Maximum | 83.00 °F |
| Minimum | 59.00 °F |

The plot shows the distribution skewed to the right, but most of the data lies within 3 standard deviations of the average value of temperature. Therefore, the data is predictable, and there is no outlier. Half of

the data lies within the short range of 74 °F – 83 °F.  
Thus, it can be assumed that Honolulu weather is warm.

A graph of a temperature

Description automatically generated

**Part 2**

Using the data analysis in Part1, a Flask API is created. A homepage is designed with all the query options as various routes, as follows.

A white text on a black background

Description automatically generated

1. The first route shows the dates and precipitation for each day collected at all stations combined from 2016-08-23 to 2017-08-23. The query result is a dictionary in JSON representation as shown:

A screenshot of a computer

Description automatically generated

1. The second route shows a list of all weather stations in Honolulu County.

A computer screen shot of a number

Description automatically generated

1. This route, displays all recorded temperatures at the weather station that has most actively collected data over the past one year (2016-8-18 to 2017-8-18). This is the data used to plot the histogram in Part 1.

A group of people in different colors

Description automatically generated

1. For the fourth and the fifth route options, one common function is used for both. For the 4th route, a start date can be specified, but the end date is the default value of 2017-08-23. For the last option, both the start and the end dates can be specified within the range 2010-01-01 to 2017-08-23.

The dates can be changed on the URL, directly.

In both cases, maximum, minimum and the average temperatures in the given date-range are displayed. An example is shown below.

A screenshot of a computer

Description automatically generated

**Conclusion**:

* Part 1 is used to understand and analyze the data.
* Part 2 uses the analysis from Part 1 to create and design an API. This enables a user to make assessments regarding planning a vacation trip etc.