

Coding Challenge #1

TRABUS Technologies

1 Challenge

Every day, climate data are being stored and processed in various databases which can help scientists understand the weather trends of local communities. In this challenge, you will **analyze hourly climate data** and **visualize temperature and rainfall patterns** in Melbourne, FL and San Diego, CA.

2 Instructions

You will be using **Python 3** and **Javascript** for this challenge along with the **Django** web framework, **Pandas**, **Psycopg2** and **Plotly** packages and any interactive javascript plotting library (Chart.js or Plotly). You may use any additional packages as you wish.

2.1 Data Ingestor

You are provided hourly climate data for 2 sites - Melbourne, FL and San Diego, CA (2 csv files provided). Using Python, write a program that reads the 2 datasets and stores the data into a Postgresql database (hint: use the psycopg2 library). The database can be a local instance running on your computer or on the cloud (<http://www.elephantsql.com> provides a free postgres server). Your database should have 1 table that has the station id, a timestamp and the variables - temperature, dewpoint, wind_speed and precipitation. In your README file, clearly document, how to create the database, how to create the tables and how to load the data into the table using the csv files. There should be only 1 table with data from both stations.

2.2 Data Analysis

Read hourly climate data values from the database for the site, KMLB (Melbourne, FL). Use the Pandas library to read the data into a dataframe and using Pandas, compute the following:

1. The total number of hours that it rained using the `precip_1hr` column
2. The total number of hours when temperature was greater than 100
3. The total number of hours when temperature was less than 32
4. Average temperature for each month
5. Sum of rainfall for each month
6. Total average temperature
7. Create a new column called "H" of humidity values from the listed hourly temperature (T) and dew point (TD) values using the formula:

$$humidity = 100 * \frac{\frac{\exp(17.625 * TD)}{243.04 + TD}}{\frac{\exp(17.625 * T)}{243.04 + T}} \quad (1)$$

where \exp is the exponential function.

Secondly, read in hourly climate values from the San Diego Airport from the database. Repeat the same computations as above for San Diego.

Then, using the Plotly library (<https://plotly.com/python/>), plot the temperature and humidity values as a line graph (with time on x-axis and temperature on y-axis). Mark the minimum and maximum temperature values on the graph as shaded dots. Also mark horizontal lines for the average temperature of the year (computed earlier) and the average humidity of the year. Again, using plotly, generate 3 bar plots that help compares San Diego's monthly temperature, humidity and rainfall totals with that of Melbourne's monthly values (month on x-axis, temperature/humidity/rainfall total on y-axis).

2.3 Data Visualization

Develop a web application using the web framework Django and Bootstrap (<https://getbootstrap.com/>) that reads the climate data from the database and generates monthly averages for temperature and monthly totals for rainfall (your code from 2.1 will be useful here). The Django application should display the monthly summaries as a Bootstrap table. The web application also needs to generate a graphical display of the above monthly averages. Use an interactive plotting library (such as <https://plotly.com/javascript/react/>) to generate column graphs of monthly values for temperature and rainfall (basically 2 column graphs and each graph should have 12 columns depicting averages/totals for each month in 2019).

3 Submission

Upload your code to a **private repository** on Github and push your code incrementally and regularly (even if you feel you have not completed the task fully). When you are finished, share the repository with davids@trabus.com, ericr@trabus.com, andrews@trabus.com and cathyh@trabus.com.

3.1 Before Submission

Make sure your repository includes:

- A `requirements.txt` that lists the Python packages used
- A `README.md` that shows how to run the code. Your README should clearly document the steps needed to run the data ingestor, data analysis and run the data visualization engine

Your code should run without problems using the instructions in the `README.md`.

4 Resources

You may find the following sources useful:

- Django - <https://docs.djangoproject.com/en/4.1/intro/tutorial01/>
- Bootstrap - <https://getbootstrap.com/>
- Pandas - <https://pandas.pydata.org/docs/>
- Plotly - <https://plotly.com/python/> and <https://plotly.com/javascript/>
- Chart.js - An alternative charting library - <https://www.chartjs.org/samples/latest/>