Programming Assignment 3 - Dino Fun World Analysis
Data Visualization (CSE 578)

The administrators of Dino Fun World, a local amusement park, have asked you, one of their data analysts, to perform three data analysis tasks for their park. These tasks will involve understanding, analyzing, and graphing attendance data for three days of the park's operations that the park has provided for you to use. They have provided the data in the form of a database, described below.

Provided Database

The database provided by the park administration is formatted to be readable by any SQL database library. The course staff recommends the sqlite3 library. The database contains three tables, named 'checkin', 'attractions', and 'sequences'. The information contained in each of these tables is listed below:

checkin:

- Description: check-in data for all visitors for the day in the park. The data includes two types of check-ins, inferred and actual checkins.
- Fields: visitorID, timestamp, attraction, duration, type

attraction:

- The attractions in the park by their corresponding AttractionID, Name, Region, Category, and type. Regions are from the VAST Challenge map such as Coaster Alley, Tundra Land, etc. Categories include Thrill rides, Kiddie Rides, etc. Type is broken into Outdoor Coaster, Other Ride, Carussel, etc.
- Fields: AttractionID, Name, Region, Category, type

sequences:

- The check-in sequences of visitors. These sequences list the position of each visitor to the park every five minutes. If the visitor has not entered the part yet, the sequence has a value of 0 for that time interval. If the visitor is in the park, the sequence lists the attraction they have most recently checked in to until they check in to a new one or leave the park.
- Fields: visitorID, sequence

The database is named 'dinofunworld.db' and is in the 'read only' folder of the Jupyter Notebook environment.

Assignment

1: The park's administrators would like you to help them understand the different paths visitors take through the park and different rides they visit. In this mission, they have selected 5 visitors at random whose checkin sequences they would like you to analyze. For now, they would like you to construct a distance matrix for these 5 visitors. The five visitors have the ids: 165316, 1835254, 296394, 404385, and 448990.

2: The park's administrators would like to understand the attendance dynamics at each ride (note that not all attractions are rides). They would like to see the minimum (non-zero) attendance at each ride, the average attendance over the whole day, and the maximum attendance for each ride on a Parallel Coordinate Plot.

3: In addition to a PCP, the administrators would like to see a Scatterplot Matrix depicting the min, average, and max attendance for each ride as above.

Instructions

This assignment will be graded by Coursera's grading system. In order for your answers to be correctly registered in the system, you must place the code for your answers in the cell indicated for each question. In addition, you should submit the assignment with the output of the code in the cell's display area. The display area should contain only your answer to the question with no extraneous information or else the answer may not be picked up correctly. Each cell that is going to be graded has a set of comment lines at the beginning of the cell. These lines are extremely important and must not be modified or removed.

A correct submission would result in feedback as:"Correct!"

An incorrect submission would look like: "Incorrect Response!"

Please execute each cell in Jupyter Notebook before submitting.

```
In [1]:  # Graded Cell, PartID: NDnou
         # Question 1: What is the most popular attraction to visit in the park?
         # Notes: Your output should be the name of the attraction.
         print('Hello World')

             Hello World
```

Feedback

There are four parts in the grading and each part has one test case where total points of all parts is 4. If some part of your data is incorrect, you will get a partial score of 0.50. If the submission failed,

we will return the corresponding error messages. If the submission is correct, you will see "Correct" with 1.0 point for each part.