**ReadMe:**

**Situation:** To answer a couple questions about an amusement park operations. In order to perform the requested analysis, have been provided you with a database containing information about one day of the park's operations.

**About 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 'checkins', 'attractions', and 'sequences'. The information contained in each of these tables is listed below:

Check-ins:

- 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

Questions are divided into tasks:

**Tasks:**

1. First task:

1: What is the most popular attraction to visit in the park?

2: What ride (note that not all attractions are rides) has the longest average visit time?

3: Which Fast Food offering has the fewest visitors?

4: Compute the Skyline of number of visits and visit time for the park's ride and report the rides that appear in the Skyline.

1. Second task:

1: A Pie Chart depicting visits to thrill ride attractions.

2: A Bar Chart depicting total visits to food stalls.

3: A Line Chart depicting attendance at the newest ride, Atmosfear, over the course of the 3 days (See template for detail).

4: A Box-and-Whisker Plot depicting total visits to the park's Kiddie Rides.

1. Third Task:

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.

1. Fourth Task:

1: The park's administrators are worried about the attendance at the ride 'Atmosfear' in the data window. To assuage their fears, they have asked you to create a control chart of the total attendance at this ride. Using the data provided, create a control chart displaying the attendance, the mean, and the standard deviation bands at one and two standard deviations.

2: Some of the park's administrators are having trouble interpreting the control chart graph of 'Atmosfear' attendance, so they ask you to also provide a moving average chart of the attendance in addition to the control chart created in the previous question. In this case, they request that you use 50 samples for the size of the moving average window.

3: In order to have options concerning the graphs presented, the park's administrators also ask you to provide a 50-sample moving average window with the average computed with exponential weighting (i.e. an exponentially-weight moving average) over the same 'Atmosfear' attendance data.

1. Fifth Task:

1: This task consists of only one question, which will require you to generate a dendrogram graph. Create this dendrogram using the trajectories of the visitors with the IDs: 165316, 1835254, 296394, 404385, and 448990. If you are unsure about how to create a dendrogram, please refer to the Jupyter Notebook example which is creating a dendrogram. When performing clustering over the trajectories to inform the dendrogram, use an average distance over all points in the cluster.

**Action:**

Using SQLite3, to connect and access the database.

Using a pointer to iterate through the database and find the required data.

The above ones are common for all tasks, rest task-based actions are communicated in the code through comments ☺

**Results:**

I have pushed python notebooks with results. Feel free to connect for any correction or questions.