

Dino Fun World Assignment

Purpose

In this assignment, you will be analyzing the data from the Dino Fun World amusement park dataset. You will be querying, analyzing structured data and manipulating it to answer a few questions about the park operations.

Technical Requirements

If you choose to work on your assignment locally, you can use the following versions:

- Python 3.12
- Sqlite3
- Pandas == 1.5.3
- Matplotlib == 3.8.3
- Numpy == 1.26.4

Assignment Description

You, in your role as a data explorer and visualizer, have been asked by the administrators of a small amusement park in your hometown to answer a few questions about their park operations. The dataset that they provided for you to perform the requested analysis includes the movement and communication data captured from the park attendees' apps during one weekend (Friday, Saturday, and Sunday).

The administrators would like you to answer three relatively simple questions about the park activities on the day in question. These questions all deal with park operations and can be answered using the data provided.

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

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

Question 3: Which Fast Food offering has the fewest visitors?

Directions

Accessing Ed Lessons

You will complete and submit your work through Ed Lessons. Follow the directions to correctly access the provided workspace:

1. Go to the Canvas Assignment, "**Submission: Dino Fun World Assignment**".
2. Click the "**Load Submission...in new window**" button.
3. Once in Ed Lesson, select the assignment titled "**Dino Fun World Assignment**".
4. Review the resources provided in the demonstration.
5. When ready, click on the code challenge and start working in the notebook titled "**Assignment1.ipynb**".

Assignment Directions

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 database file is named '**dinofunworld.db**' and is available in the '**/course/data/CSE-578/dinofunworld.db**' path.

Note: Please note that the database file is accessible through the learner submission workspace, which requires establishing a connection with the database. For downloading the dataset and potentially working locally, refer to the overview document page.

The information contained in each of these tables is listed below:

checkin:

- The 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, Carousel, etc.
- Fields: AttractionID, Name, Region, Category, type

sequences:

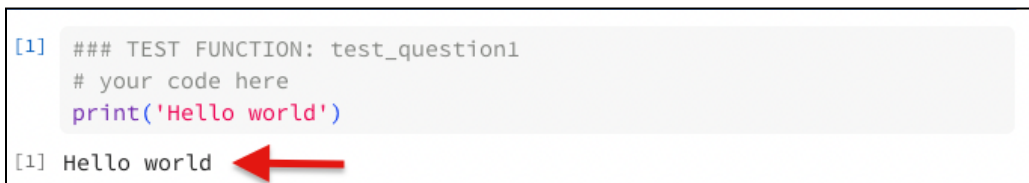
- The check-in sequences of visitors. These sequences list the position of each visitor to the park every five minutes. If the visitor has not entered the park yet, the sequence has a value of 0 for that time interval. If the visitor is in the park, the sequence lists the attraction they have most recently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

Using the provided data, answer the three questions that the administrators have asked.

Submission Directions for Assignment Deliverables

This assignment will be auto-graded. You must complete and submit your work through Ed Lesson's code challenge to receive credit for the course:

1. 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.
 - a. 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.
 - b. Each cell that is going to be graded has a set of comment lines (ex: `### TEST FUNCTION: test_question1`) at the beginning of the cell. **This line is extremely important and must not be modified or removed.**
2. After completing the notebook, run each code cell individually or click "**Run All**" at the top to print the outputs.



The screenshot shows a Jupyter Notebook cell. The code area contains three lines: `### TEST FUNCTION: test_question1`, `# your code here`, and `print('Hello world')`. Below the code area, the output area displays `[1] Hello world`. A red arrow points from the output text to the left, indicating the output of the code cell.

3. When you are ready to submit your completed work, click on "**Mark**" at the bottom right of the screen.
4. You will know you have successfully completed the assignment when feedback appears for each test case with a score.
5. If needed: to resubmit the assignment in Ed Lesson
 - a. Edit your work in the notebook
 - b. Run the code cells again
 - c. Click "**Mark**" at the bottom of the screen

Your submission will be reviewed by the course team and then, after the due date has passed, your score will be populated from Ed Lesson into your Canvas grade.

Evaluation

There are three parts in the grading, and each part has one test case where the total number of points for all parts is 50. If the submission fails, we will return the corresponding error messages. If the submission is correct, you will see "Correct" with scores for each part.