**Analytics Test**

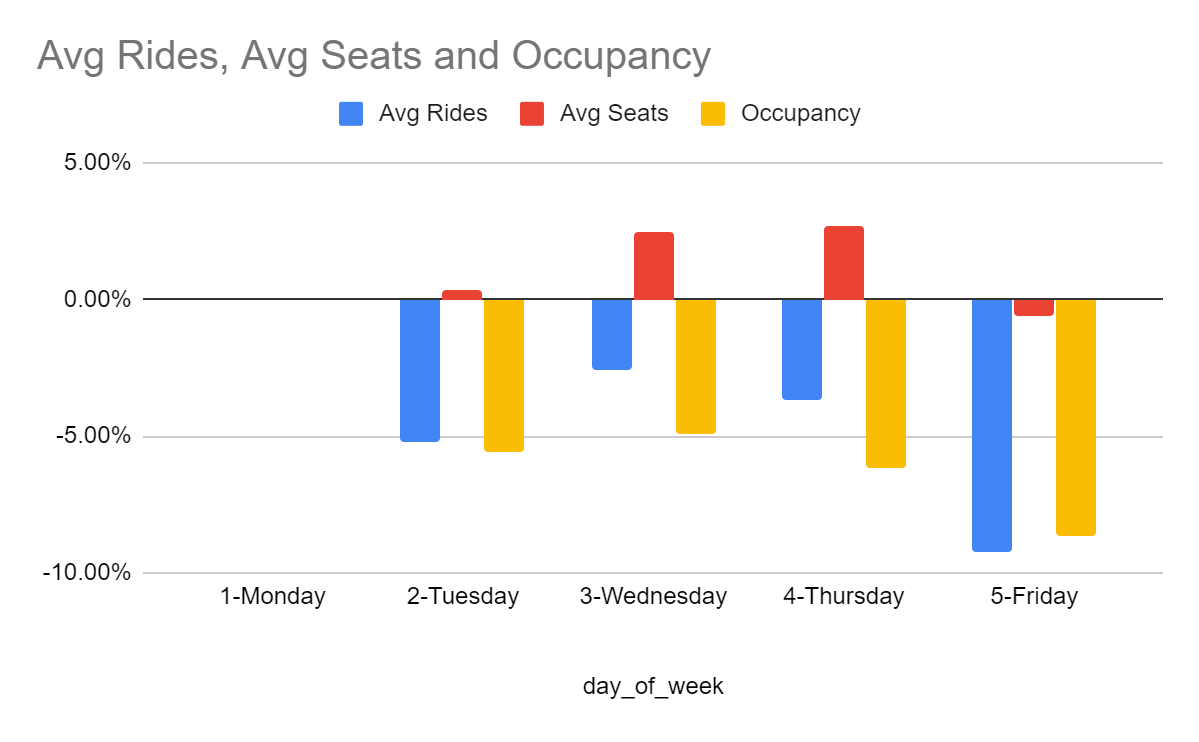
Dataset Link for Questions 2-4: [Data Set - BA/DA Test](https://docs.google.com/spreadsheets/d/1PU4rGJjLX1xUJmnEWWdd7MraiNNCXScAAtY-TA9sYuY/edit?usp=sharing)

**Question 1. Explore insights/trends/patterns from the below graph.**

Below are the week of day pattern on average rides taken by customers, average seats operated by Cityflo and average occupancy (fill rate of buses: rides/seats) over the last 3 months. Each of these values is plotted as deviation from the value on Monday.

Find insights based on this graph, and suggest measures to improve fleet scheduling. Should all buses be run everyday? Should all routes and timings be run everyday?

Note: All values are on the base of Monday.



**Question 2. Calculate the weekly streak value of customers**

We have the information of the customer and each ride they have taken to date. We want to find the running streak of each customer, in each calendar week

Write code to calculate the customer's weekly streak (Travelled at least once in the week), given the customer ride table as input table and show the result in the format of the sample output set.

Below is the sample input table - **Sample Customer Ride Table** with customer name and the ride date.

Customer Ride Data:

|  |  |
| --- | --- |
| **Customer Name** | **Ride Date** |
| Natasha | 09-01-2020 |
| Natasha | 13-01-2020 |
| Natasha | 15-01-2020 |
| Natasha | 17-01-2020 |
| Natasha | 20-01-2020 |
| Natasha | 22-01-2020 |
| Natasha | 24-01-2020 |
| Natasha | 28-01-2020 |
| Natasha | 30-01-2020 |
| Natasha | 30-01-2020 |
| Natasha | 31-01-2020 |
| Natasha | 17-02-2020 |
| Natasha | 18-02-2020 |
| Karan | 03-01-2020 |
| Karan | 17-02-2020 |
| Karan | 18-02-2020 |
| Karan | 20-02-2020 |
| Karan | 24-02-2020 |

**Sample Output table**

For each calendar week (Week 1 is 1st week of Jan 2020 - ending on 1st Sunday in January), find the streak value for each customer.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Week1** | **Week2** | **Week3** | **Week4** | **Week5** | **Week6** | **Week7** | **Week8** | **Week9** |
| **Natasha** | - | New Customer | 1 | 2 | 3 | -1 | -2 | 1 | -1 |
| **Karan** | New Customer | -1 | -2 | -3 | -4 | -5 | -6 | 1 | 2 |

Values:

If the customer has used Cityflo this week, the streak value changes by +1

If the streak is broken, and the customer doesn’t use Cityflo this week, the streak value resets to -1.

If the customer has not used Cityflo this week, the streak value changes by -1.

If the customer restarts using Cityflo this week, the streak value, resets to +1

+ive value is weekly streak with more than 0 rides in consecutive weeks.

-ive value is weekly streak without t aking ride for n weeks

**Weekn** : Nth week of year starting from 1st Jan, 2020.

**New Customer:** Customer’s first ride with Cityflo

You can solve this in SQL, R, Python or any other language you are comfortable with.

Please provide executable code. Please add comments to explain each of your steps.

**Question 3. Identify stops from customers locations**

Given below is the dataset that consists of customer id and their home location coordinates. Write an algorithm that suggests Cityflo bus stops, such that maximum customers are served with minimal walking distance from their home to the Cityflo bus stop. Each additional stop leads to an increase in the bus trip time, therefore minimise the number of stops as far as possible.

|  |  |  |
| --- | --- | --- |
| **customer\_id** | **home\_location\_lat** | **home\_location\_long** |
| 194861 | 19.23153992 | 72.97229655 |
| 161762 | 19.2130542 | 72.9541699 |
| 193540 | 19.2621575 | 72.9670945 |
| 138047 | 19.231485 | 72.970033 |
| 192559 | 19.22796885 | 72.96774686 |

Full Dataset link: [Data Set - BA/DA Test](https://docs.google.com/spreadsheets/d/1PU4rGJjLX1xUJmnEWWdd7MraiNNCXScAAtY-TA9sYuY/edit#gid=798864292)

Please highlight your approach for solving this problem as steps.

(Optional): Provide code in SQL, Python, R or simple psuedo code.

**Question 4. Identify routes from suggested locations**

File Link: [Data Set - BA/DA Test](https://docs.google.com/spreadsheets/d/1PU4rGJjLX1xUJmnEWWdd7MraiNNCXScAAtY-TA9sYuY/edit#gid=868645309)

Below is the table of customers who have suggested new routes for Cityflo to start. Dataset contains customer id and their home location and office location coordinates. Create an algorithm to identify different home-to-office routes which cater to maximum number of customers. Each route has to complete all the pickups before initiating any of the drops. Each route can have multiple pickup and drop points, but should have minimised travel time.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **customer\_id** | **home\_location\_lat** | **home\_location\_long** | **office\_location\_lat** | **office\_location\_long** |
| 186262 | 19.0614318 | 72.8677735 | 19.2136775 | 72.9903447 |
| 130699 | 19.0687681 | 72.8778164 | 19.1662566 | 72.8525696 |
| 195493 | 19.281369 | 72.858099 | 19.1167474 | 72.8631927 |
| 195479 | 19.235013 | 73.0131951 | 18.9255728 | 72.8242221 |
| 11094 | 19.1729281 | 72.8359056 | 19.006396 | 72.833292 |
| 80504 | 19.1135684 | 72.8654484 | 19.2119231 | 72.8680391 |
| 64523 | 19.184193 | 72.965165 | 19.006287 | 72.829433 |

Please highlight your approach for solving this problem as steps.

(Optional): Provide code in SQL, Python, R or simple psuedo code.

**Question 5 :** Current Fare is Rs F, and the no. of active customers is C. An X% increase in Fare, results in Y% decrease in customers. Find the relation between X & Y such that, new revenue is greater than current revenue.

Details:

1. One customer takes one ride a day
2. Revenue = No of Customers \* Average Fare