ASTRA

satu Indonesia
Semangat Astra Terpadu Untuk Indonesia

Astra Data Scientist Bootcamp 3
(Technical Section)

AstraDigital

Rules

1. Technical Section consist of SQL, Python, and Statistics.
2. You have 50 minutes to solve the technical problems and submit your answers.
3. You can use any references, but you are not allowed to work with other participants.
4. Write your answers for technical questions in presentation template (PowerPoint, Google Slides, etc.).
5. Export to PDF and name the file as: “[YOUR TEST NUMBER]_[YOUR NAME]_TECHNICAL.pdf“.
6. Send your submission using your mail registered in *astra.virtue* to https://forms.gle/8j6YM6ei4QGN2gdC9 before 08.50 AM.
7. Follow the rules. We will take notes on participants who breaks the rules.
8. Read and answer the questions carefully.

SQL

Given a database called `bigquery-public-data`.

From the database, you have dataset `new_york_citibike` that have two tables which are:

- `citibike_stations`
- 'citibike_trips`

** Notes:
Those tables are in separated excel called 'Sample SQL Table.xlsx' which you can find in gdrive.

Questions

1. Show top 5 stations that have the most capacity, sort it from high to low **(3 marks).**
2. Show number of trips generated in 2018 – 2020 grouping by gender for each station in questions 1 as a start station **(7 marks).**

Write your full queries and mention the languages used to solve the question i.e SQLite, PostGreSQL, etc.

Python

Suppose you have a loaded dataset in python environment called `data_waktu`.

The dataset has 365 rows and a column called `waktu`, a string value of date and time with format '%m-%d-%Y %H:%M:%S' .

Questions

1. Create column called `random_int` that has value from 0 to 5 (included) using `random` packages/library in python. **(2 marks)**
2. From column `waktu`, create new column called `nama_hari` that show the name of days (Monday, Tuesday, Wednesday, etc.) **(4 marks)**
3. Create new table that sum column `random_int` for each `nama_hari` in the table. **(4 marks)**

Sample output for python question 1-2

```
1  data_waktu.sample(7, random_state=2021)
```

executed in 14ms, finished 09:09:23 2021-01-15

	waktu	random_int	nama_hari
250	09-23-2020 09:03:28	2	Wednesday
15	02-01-2020 09:03:28	1	Saturday
0	01-17-2020 09:03:28	1	Friday
113	05-09-2020 09:03:28	5	Saturday
22	02-08-2020 09:03:28	3	Saturday
124	05-20-2020 09:03:28	0	Wednesday
312	11-24-2020 09:03:28	1	Tuesday

Sample output for python question 3

	random_int
nama_hari	
Friday	133
Monday	120
Saturday	128
Sunday	147
Thursday	134
Tuesday	109
Wednesday	140

Write your full codes in python to solve the questions.

Statistics

As a Data Analyst, you want to tell Marketing Team about the effects of two different marketing programs.
The first program is using funny concept (first group) and the second one is using elegant concept (second group).
The program is sent to 10 customers for each group.
After the program ends you calculated the spending of each customers.

Given:

- alpha = 5%
- average spending of first group = 100
- average spending of second group = 95
- pooled standard deviation = 4.3

Questions:

1. Calculate t-statistics. **(4 marks)**
2. Prove whether different program have a significant different effect on customer's spending. **(6 marks)**

Write the whole calculation (not programming code, but your manual step calculations), not only your final answer.

Astra Data Scientist Bootcamp 3

Good Luck !