**TASK- 1**

Questions;

For the trips between 1st of October 2022 and 7th of October 2022;

1. What is the average trip\_miles, average trip\_seconds, average fare, and average tips?

SELECT

  AVG(trip\_miles) AS average\_trip\_miles,

  AVG(trip\_seconds) AS average\_trip\_seconds,

  AVG(fare) AS average\_fare,

  AVG(tips) AS average\_tips

FROM bigquery-public-data.chicago\_taxi\_trips.taxi\_trips

WHERE trip\_start\_timestamp >= '2022-10-01 00:00:00' AND trip\_end\_timestamp <= '2022-10-07 23:59:59';

| Row | average\_trip\_miles | average\_trip\_seconds | average\_fare | average\_tips |  |
| --- | --- | --- | --- | --- | --- |
| 1 | 6.5554199683042764 | 1294.3901665344961 | 22.076571428571437 | 1.99224603174603 |  |

1. How many trips had a bigger tip amount than the fare?

SELECT COUNT(\*) AS trips\_with\_bigger\_tip

FROM bigquery-public-data.chicago\_taxi\_trips.taxi\_trips

WHERE tips > fare AND trip\_start\_timestamp >= '2022-10-01 00:00:00' AND trip\_end\_timestamp <= '2022-10-07 23:59:59';

| Row | trips\_with\_bigger\_tip |  |
| --- | --- | --- |
| 1 | 4 |  |

1. What is the payment\_type distribution based on the trip counts?

SELECT payment\_type, COUNT(\*) AS trip\_count

FROM bigquery-public-data.chicago\_taxi\_trips.taxi\_trips

WHERE trip\_start\_timestamp >= '2022-10-01 00:00:00' AND trip\_end\_timestamp <= '2022-10-07 23:59:59'

GROUP BY payment\_type;

| w | payment\_type | trip\_count |  |
| --- | --- | --- | --- |
| 1 | Credit Card | 300 |  |
| 2 | Dispute | 1 |  |
| 3 | Cash | 483 |  |
| 4 | No Charge | 1 |  |
| 5 | Unknown | 105 |  |
| 6 | Prcard | 199 |  |
| 7 | Mobile | 173 |  |

1. Which company has the most trips?

SELECT company, COUNT(\*) AS trip\_count

FROM bigquery-public-data.chicago\_taxi\_trips.taxi\_trips

WHERE trip\_start\_timestamp >= '2022-10-01 00:00:00' AND trip\_end\_timestamp <= '2022-10-07 23:59:59'

GROUP BY company

ORDER BY trip\_count DESC

LIMIT 1;

| Row | company | trip\_count |  |
| --- | --- | --- | --- |
| 1 | Flash Cab | 387 |  |

5- What insights can you collect from the dataset about the morning and evening trips? A morning

trip is between 6 and 9 am and an evening trip is between 4 and 7 pm.

SELECT COUNT(\*) AS morning\_trip\_count, AVG(trip\_miles) AS morning\_average\_trip\_miles, AVG(fare) AS morning\_average\_fare

FROM bigquery-public-data.chicago\_taxi\_trips.taxi\_trips

WHERE EXTRACT(HOUR FROM trip\_start\_timestamp) >= 6 AND EXTRACT(HOUR FROM trip\_start\_timestamp) <= 9

AND trip\_start\_timestamp >= '2022-10-01 00:00:00' AND trip\_end\_timestamp <= '2022-10-07 23:59:59';

| Row | morning\_trip\_count | morning\_average\_trip\_miles | morning\_average\_fare |  |
| --- | --- | --- | --- | --- |
| 1 | 189 | 6.1833862433862414 | 20.907883597883604 |  |

SELECT COUNT(\*) AS evening\_trip\_count, AVG(trip\_miles) AS evening\_average\_trip\_miles, AVG(fare) AS evening\_average\_fare

FROM bigquery-public-data.chicago\_taxi\_trips.taxi\_trips

WHERE EXTRACT(HOUR FROM trip\_start\_timestamp) >= 16 AND EXTRACT(HOUR FROM trip\_start\_timestamp) <= 19

AND trip\_start\_timestamp >= '2022-10-01 00:00:00' AND trip\_End\_timestamp <= '2022-10-07 23:59:59';

| Row | evening\_trip\_count | evening\_average\_trip\_miles | evening\_average\_fare |  |
| --- | --- | --- | --- | --- |
| 1 | 330 | 6.3435151515151507 | 21.9529393939393 |  |

**TASK 2**

PYTHON SCRIPT

import tweepy

# Twitter API credentials

consumer\_key = "ETOf6Va7Q2fTstMnyK4uthigB"

consumer\_secret = "f9sv1PgqcCXkBbVsrr4qEOBOOzMpm19aMDUe1leeDne9KbYOi4"

access\_key = "1664341244539969559-zDWtSLFycVQTgvC3gUKGambSi2JokO"

access\_secret = "Nc7pqqIyRihyaTcHxYlWTxRJsdsjUGNmPLXFsUdTpy10H"

# Authorize Twitter and initialize Tweepy

auth = tweepy.OAuthHandler(consumer\_key, consumer\_secret)

auth.set\_access\_token(access\_key, access\_secret)

api = tweepy.API(auth)

print(api)

<tweepy.api.API object at 0x7f37be5ce680>

# Scrape tweets from the "BuildingMindsAI" Twitter account

screen\_name = "BuildingMindsAI"

tweet\_count = 1000  # Number of tweets to scrape

tweets = []

try:

    for tweet in tweepy.Cursor(api.user\_timeline, screen\_name=screen\_name, tweet\_mode="extended").items(tweet\_count):

        tweets.append(tweet.full\_text)

except tweepy.errors.TweepyException as e:

    print("Error: " + str(e))

# Save the tweets in a text file

filename = "BuildingMindsAI\_tweets.txt"

with open(filename, "w", encoding="utf-8") as file:

    file.write("\n".join(tweets))

print(f"{len(tweets)} tweets scraped and saved to {filename}.")

Error: 403 Forbidden

453 - You currently have access to a subset of Twitter API v2 endpoints and limited v1.1 endpoints (e.g. media post, oauth) only. If you need access to this endpoint, you may need a different access level. You can learn more here: <https://developer.twitter.com/en/portal/product>

0 tweets scraped and saved to BuildingMindsAI\_tweets.txt

**AS there are limited access twitter has changed their(Policies) elevated research access.Its paid version now. Previously it was there for free.**

**Some another approach**

import requests

from bs4 import BeautifulSoup

def scrape\_tweets(BuildingMindsAI):

    url = f"https://twitter.com/{screen\_name}"

    response = requests.get(url)

    soup = BeautifulSoup(response.content, "html.parser")

    tweets = soup.find\_all("div", class\_="tweet")

    with open(f"{screen\_name}\_tweets.csv", "w", encoding="utf-8") as file:

        for tweet in tweets:

            text = tweet.find("div", class\_="tweet-text").get\_text().strip()

            file.write(text + "\n")

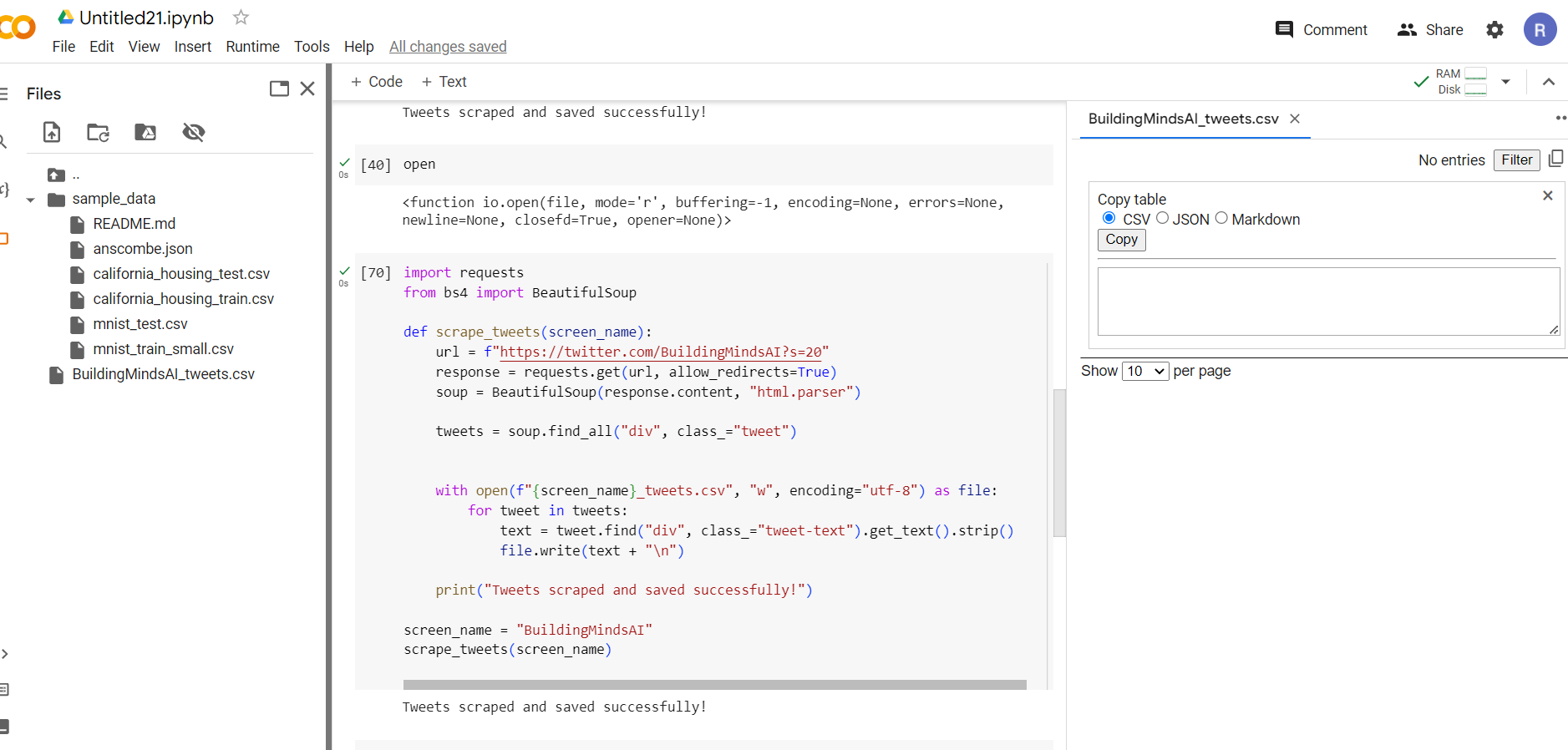
    print("Tweets scraped and saved successfully!")

screen\_name = "BuildingMindsAI"

scrape\_tweets(screen\_name)

Tweets scraped and saved successfully!

**Text file is generated but, tweets were un-abled to scrape, due to twitter changed (endpoint access) policies.**

****