Task #1 Data Analysis Note: Analysis the Bike-share.

Name: Ruoyi Jin

Date:7/17/2022

Data from: Cyclistic (fictional company)

There are six steps for me to do this project:

Ask, Prepare, Process, Analysis, Share, Act.

Ask: The problem I am trying to solve is: How does a bike_share Navigate is Speedy Success.

Prepare: The data is downloaded, and it is located from the google data share.

The data is organized and there is no bias.

The data is ROCCC. (reliable, original, comprehensive, current, cite)

Sort the data by range: started time and User_ID

Error Found:

1. [202004-divvy-tripdata] By using the conditional formattiing to check the spreadsheet, the started time and ended time is entry wrong.

   For example: In the spreadsheet, the start time is 2020-04-07 7:53:54, and the ended time is 2020-04-07 7:53:46.

   Found empty cells and delected them, decrease the number of cell from 84777 to 84677
2. [202005-divvy-tripdata] By using the conditional formattiing to check the spreadsheet,. The started time and ended time is entry wrong.

   For example: In the spreadsheet, the start time is 5/1/2020 10:03:49, and the ended time is 5/1/2020 10:03:41.

   Found empty cells and delected them, decrease the number of cell from 200275 to 199954

   Found user_id is 0 and delected, decrease the number of cell from 199954 to 199952.
3. [202006-divvy-tripdata]By using the conditional formattiing to check the spreadsheet,. The started time and ended time is entry wrong.

   For example: In the spreadsheet, the start time is 6/3/2020 14:03:49, and the ended time is 6/3/2020 14:03:41

   Found user_id is 0 and delected, decrease the number of cell from 343006 to 343005.

   Found empty cells and delected them, decrease the number of cell from 343005 to 342536

Process: Creat a new column named “ride_length” apply the calculation:D2-C2

Creat a new column named “day_of_week” apply the function =WEEKDAY(C2,1)

Analysis: Calculate the average of ride_length by applying the ufnction =AVERAGE().

Calculate the maximum of right_length by applying function =MAX().

Calculate the mode of day_of_way by applying =MODE().

Create a pivot table to show the the day_of_week, member-casual, and count the ride_id to calculate different types of users use bike in different weekday.

By viewing the data of pivot table of 2020_4, there are 17897 people used the bike on Sunday, and the longest of using bike is on Friday, the average is 0:41:33.

By viewing the data of pivot table of 2020_5, there are 46 people used the bike on Saturday, and the longest of using bike is on Sunday, the average is 0:38:25.

By viewing the data of pivot table of 2020_6, there are 46 people used the bike on Sunday, and the longest of using bike is on Sunday, the average is 0:38:18.

Observation from pivot table: By observing the pivot table, we can have in 2020, April, the total user is 84677, and the total annual user is 61089, casual users are 23588. The average ride length of using bike is 00:35:35.

By observing the pivot table, we can have in 2020, May, the total user is 19952, and the total annual user is 113189, casual users are 86763. The average ride length of using bike is 00:33:01.

By observing the pivot table, we can have in 2020, June, the total user is 342536, and the total annual user is 188027, casual users are 154509. The average ride length of using bike is 00:33:15.

Observation from SQL: By using SQL we know there are 50 stations were used in 2020, April.

Then, let us analysis the data of 2020, April, the most popular station is Clark St & Elm St, the time of using that station is 850.

Go to data of 2020, May, the most popular station is Clark St & Elm St, the time of using that station is 1941.

For data of 2020, June, the most popular station is Streeter Dr & Grand Ave, the time of using that station is 4003.

```sql
SELECT
  COUNT(April.ride_id) AS number_of_member_April,
  (
    SELECT
      COUNT(May.ride_id)
    FROM
      `2020_Q2_tripdata.May_tripdata` AS May
    WHERE
      member_casual = "member"
  )
    AS number_of_member_May,
  (
      SELECT
        COUNT(June.ride_id)
    FROM
        `2020_Q2_tripdata.June_tripdata` AS June
    WHERE
    member_casual = "member"

  )
  AS number_of_member_June
FROM
  `2020_Q2_tripdata.April-tripdata` AS April
WHERE
  member_casual = "member"
```

Query results

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS

ow	number_of_member_April	number_of_member_May	number_of_member_June
1	61089	113189	188027

Here is the figure for the top 10 popular station.

Row	the_times_of_using_station	start_station_name
1	850	Clark St & Elm St
2	730	Dearborn St & Erie St
3	719	Desplaines St & Kinzie St
4	686	St. Clair St & Erie St
5	625	Clark St & Armitage Ave
6	614	Wabash Ave & Grand Ave
7	605	Broadway & Barry Ave
8	584	Stockton Dr & Wrightwood Ave
9	576	Larrabee St & Webster Ave
10	574	Clark St & Schiller St

Row	the_times_of_using_station	start_station_name
1	1941	Clark St & Elm St
2	1558	Indiana Ave & Roosevelt Rd
3	1549	Larrabee St & Webster Ave
4	1480	Dearborn St & Erie St
5	1474	Clark St & Lincoln Ave
6	1442	Stockton Dr & Wrightwood Ave
7	1413	Dearborn Pkwy & Delaware Pl
8	1405	Clark St & Armitage Ave
9	1399	Wells St & Huron St
10	1377	Clark St & Schiller St

Row	the_times_of_using_station	start_station_name
1	4003	Streeter Dr & Grand Ave
2	3344	Clark St & Elm St
3	3023	Lake Shore Dr & North Blvd
4	2688	Indiana Ave & Roosevelt Rd
5	2641	Broadway & Barry Ave
6	2594	Wells St & Concord Ln
7	2555	Lake Shore Dr & Monroe St
8	2539	Clark St & Lincoln Ave
9	2485	Wells St & Elm St
10	2458	Larrabee St & Webster Ave