Dear Naveen,

Thank you for providing us with the two datasets from Uber Technologies Inc. The below data highlights the summary statistics from the two datasets received. Please let us know if the figures are not aligned with your understanding.

1. Data Analysis:

There are two data sets given by Uber Technologies Inc. They are **Fact\_trip** and **Dim\_city**.

1. **Fact\_trip** - It provides all the details of trip transactions.

It has the following columns.

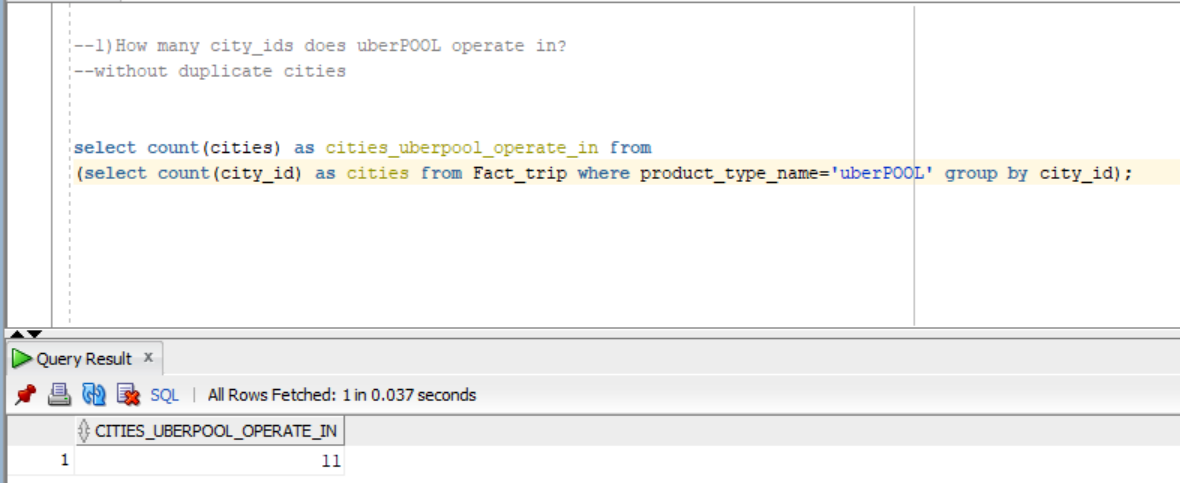
* + trip\_uuid- It is the unique id of trip.
  + datestr- It is the date of trip
  + product\_type\_name- type of services provided in trip such as Delivery,Helium,uberPOOL,uberX)
  + city\_id- It is the id of city where service is delivered to users
  + driver\_uuid- It is the unique id of driver
  + is\_completed- For successful completion of trip it shows True if not it shows False
  + eta- Estimated time of arrival
  + ata- Actual time of arrival
  + ufp\_fare – Universal fare price
  + fare\_final- It is the final fare price of travel

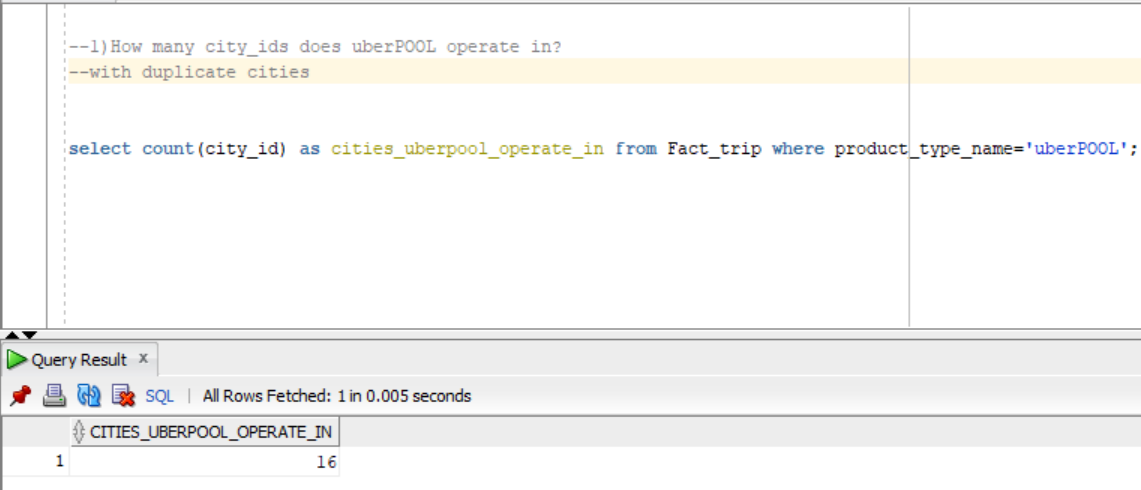
1. **Dim\_city** It provides all the details of service city and country details.

It has the following columns.

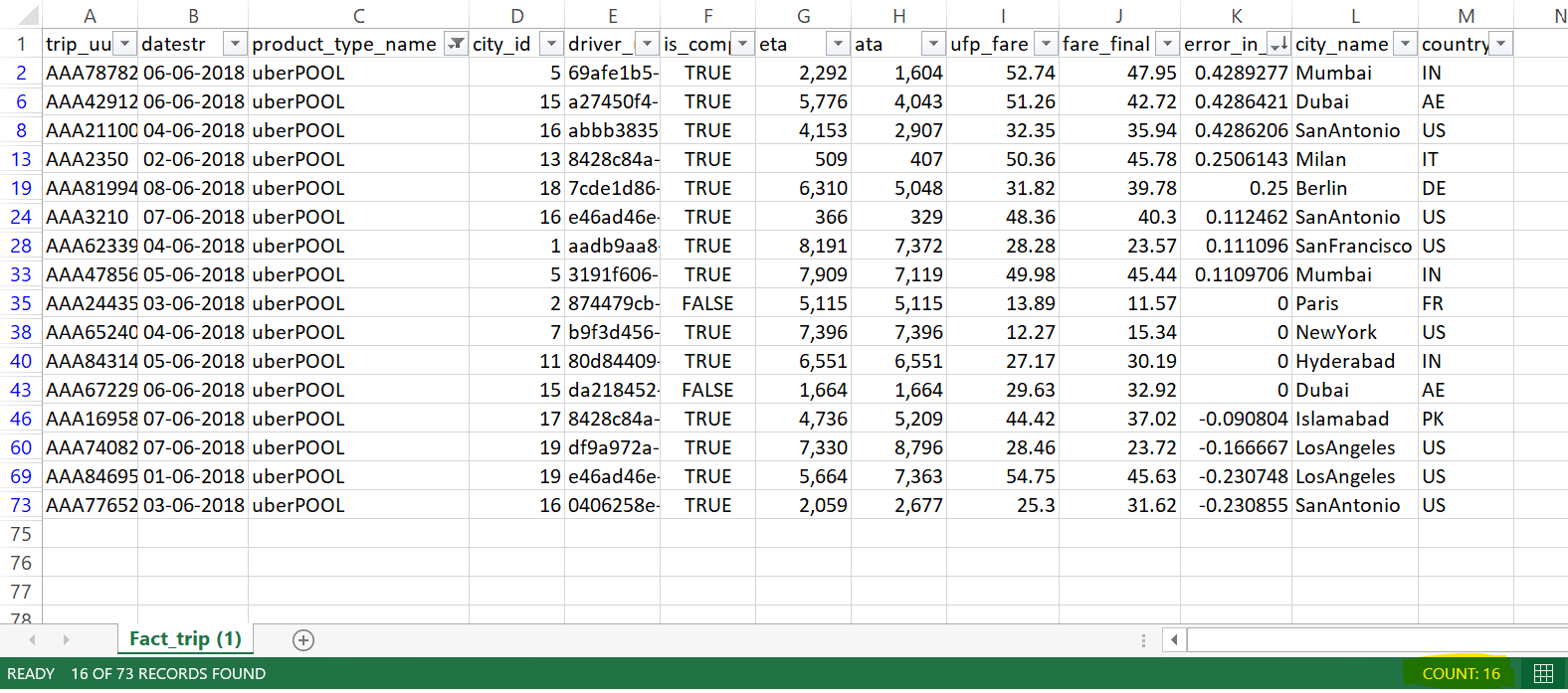
* city\_id- It is the id of city where service is delivered to users
* city\_name- It is the name of city where service has been delivered.
* Country- It is the country of the city where service has been delivered.

1. Answer the following questions:
2. How many city\_ids does uberPOOL operate in?

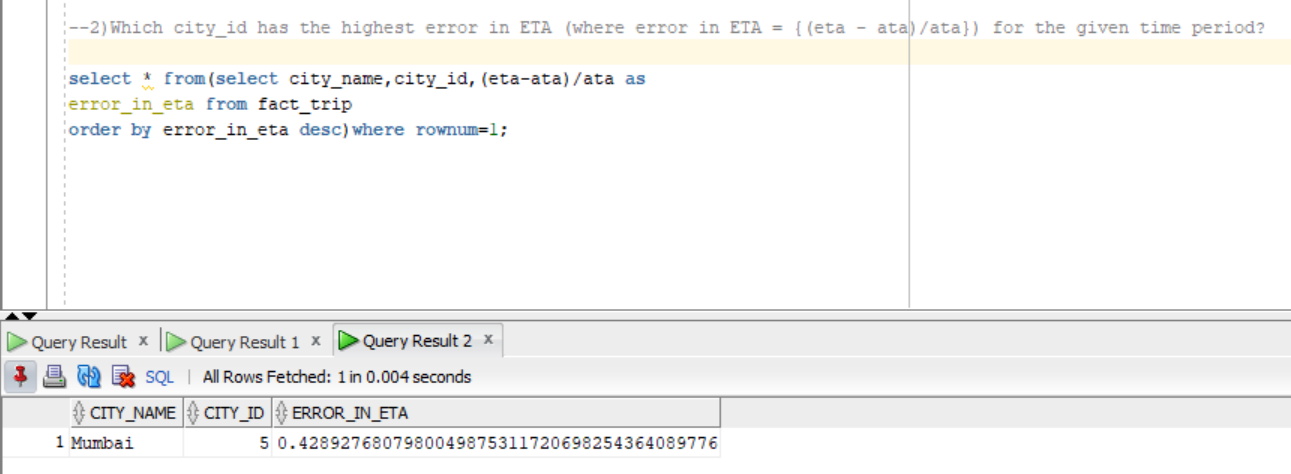




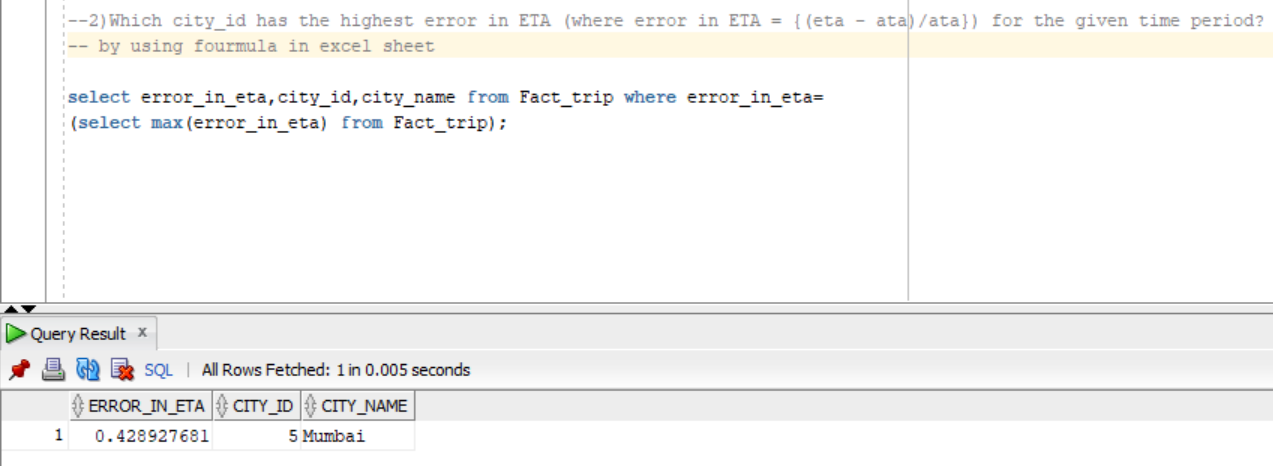
By analyzing Excel sheet we can see given snip as answer with duplicate values with 16 cities



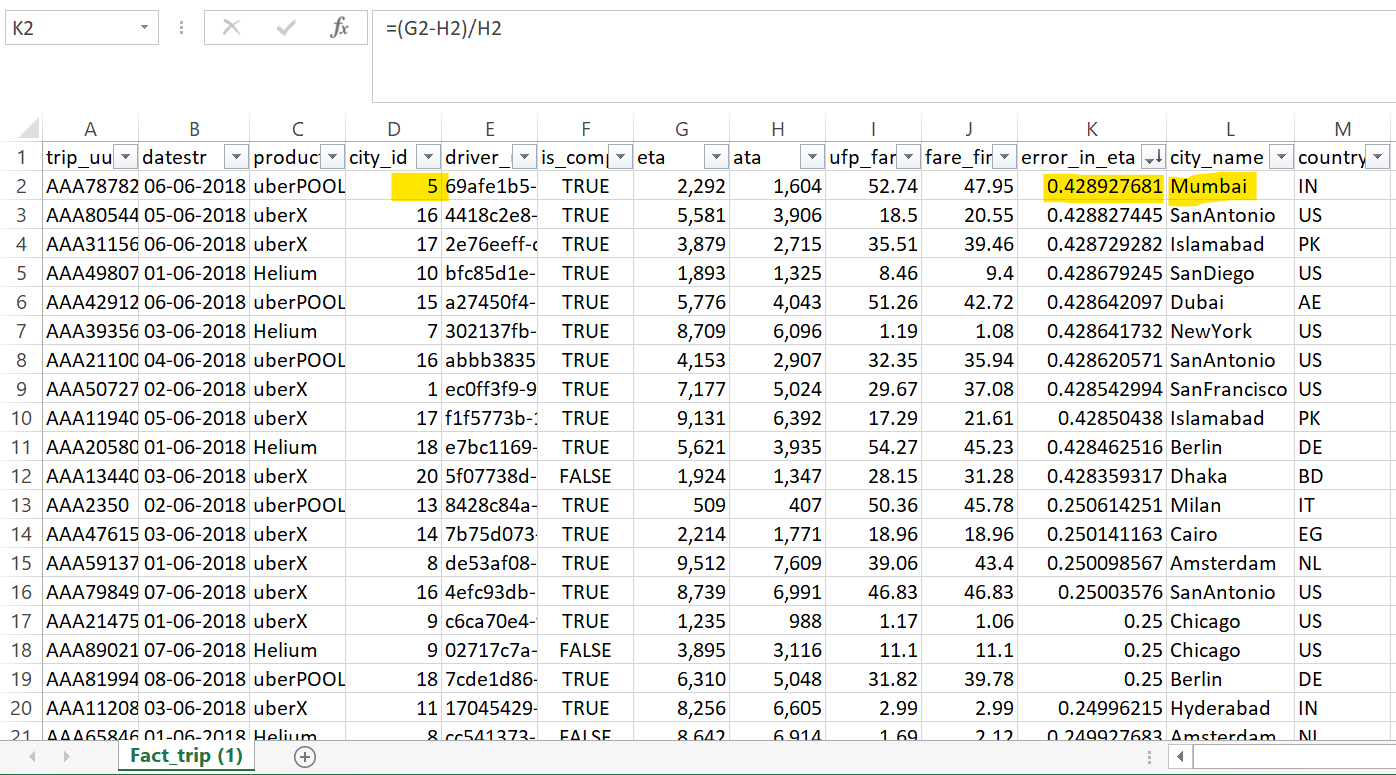
1. Which city\_id has the highest error in ETA (where error in ETA = {(eta - ata)/ata}) for the given time period?



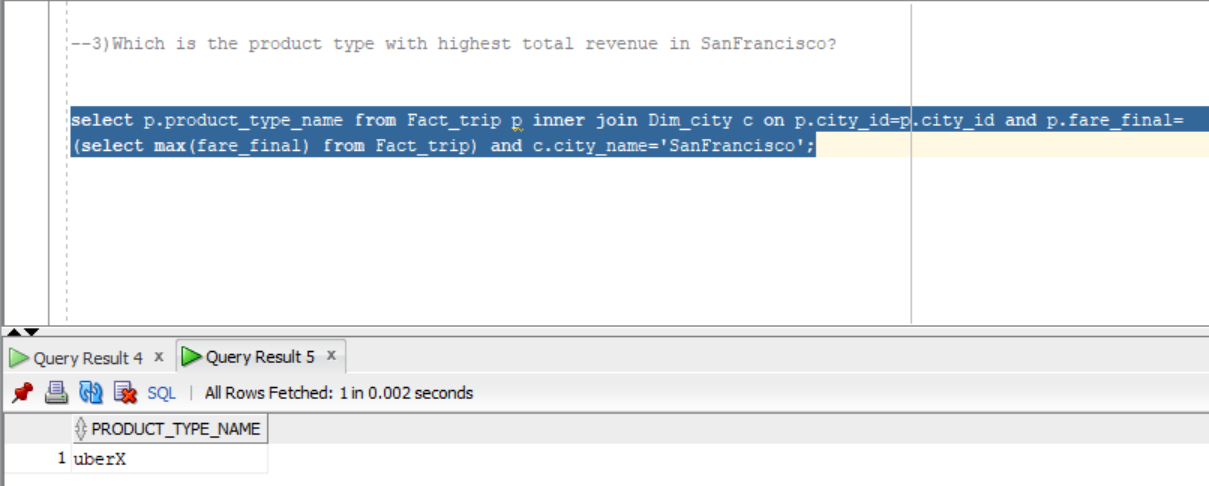
By using error in ETA = {(eta - ata)/ata}) formula in excel sheet



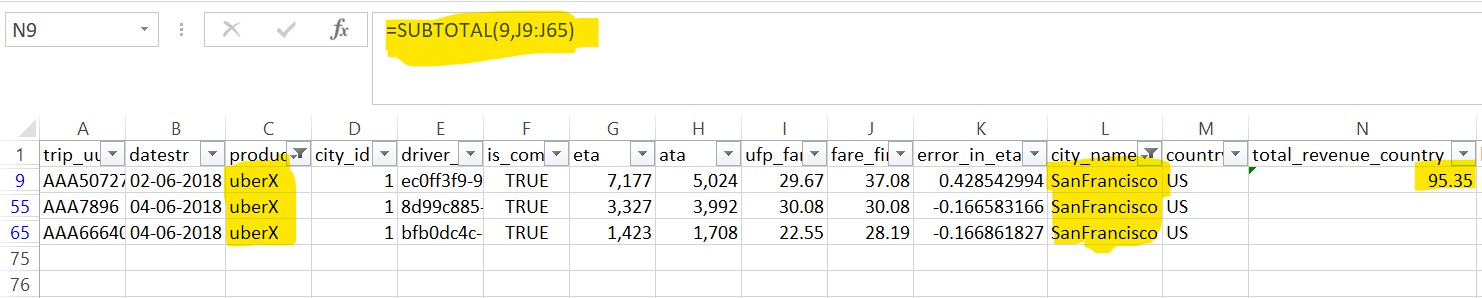
By analyzing Excel sheet we can see given snip as answer Mumbai has the highest error in ETA with 0.428927681



1. Which is the product type with highest total revenue in San Francisco?



By analyzing Excel sheet we can see given snip as answer where uberX has highest total revenue(fare\_final) in San Francisco with 95.35



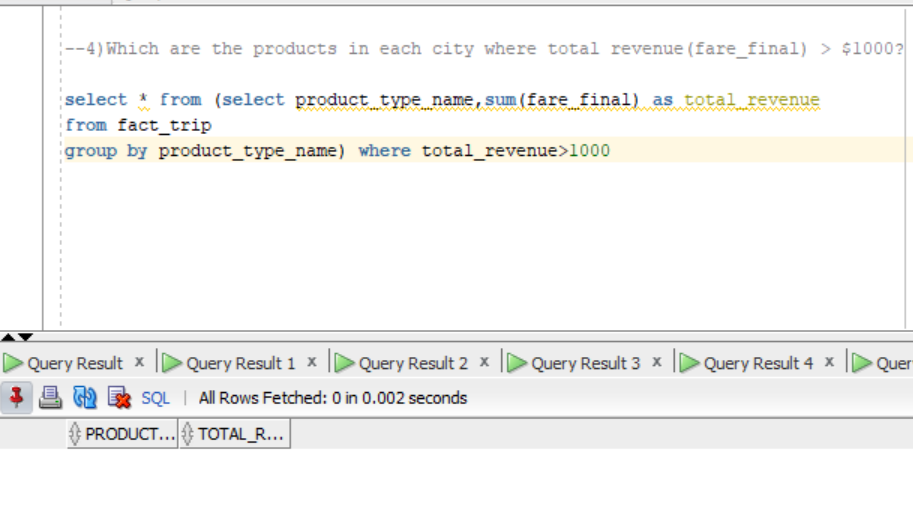
1. Which are the products in each city where total revenue (fare\_final) > $1000?

There is no output when the total fare final of each product is as given below

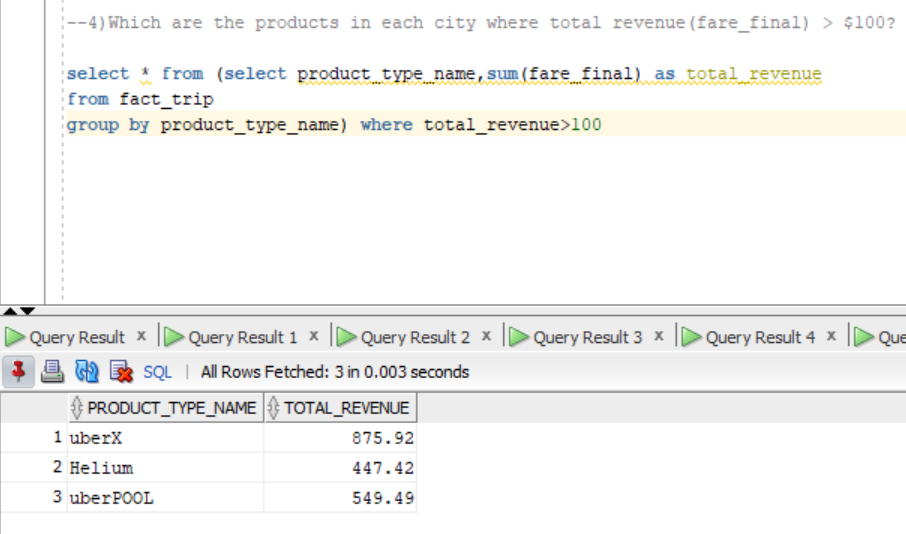
uberX - 875.92

Helium - 447.42

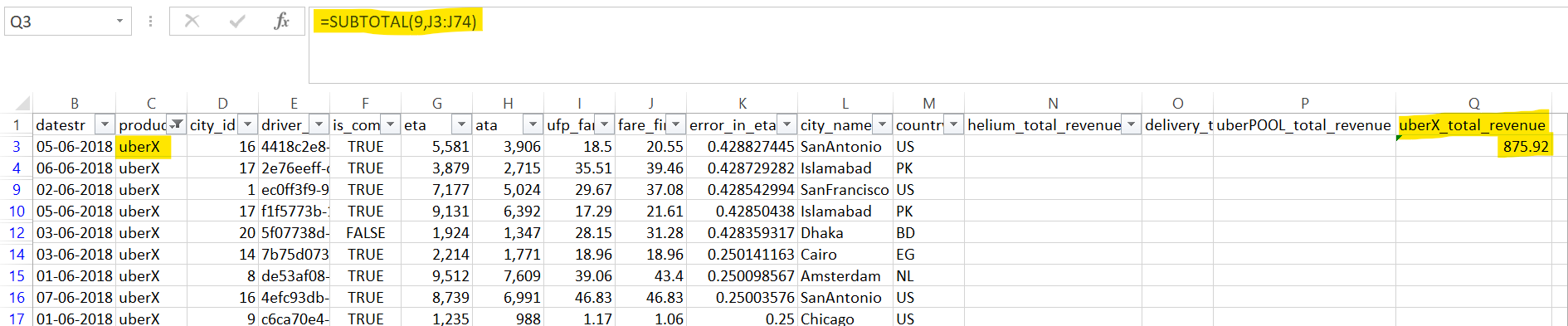
uberPOOL - 549.49

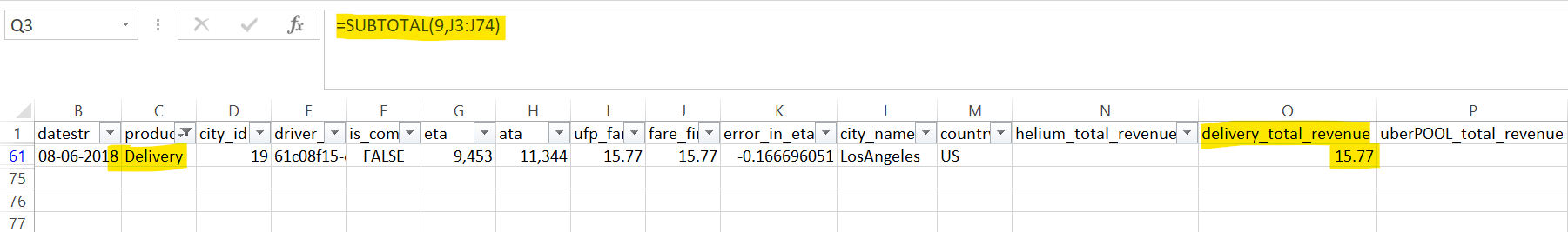


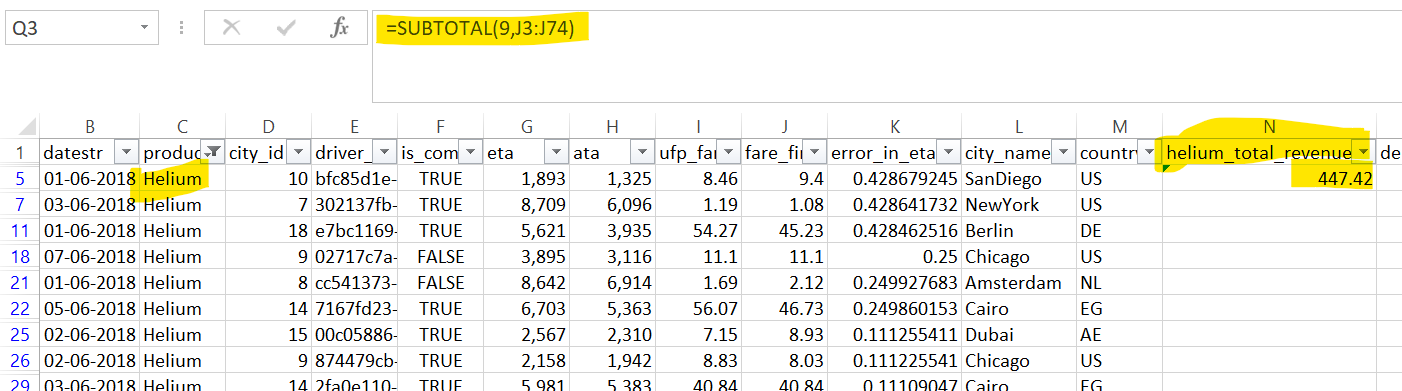
When we decrease the total revenue>100 then we can see below snip as output

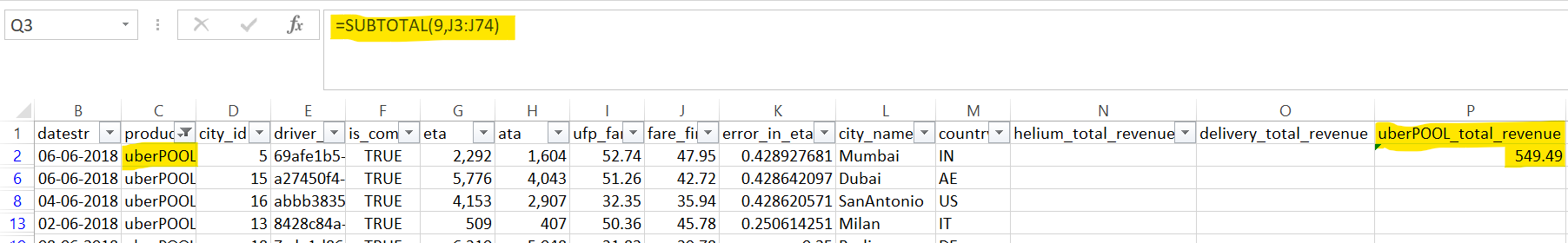


By analyzing Excel sheet we can see given snips as answers

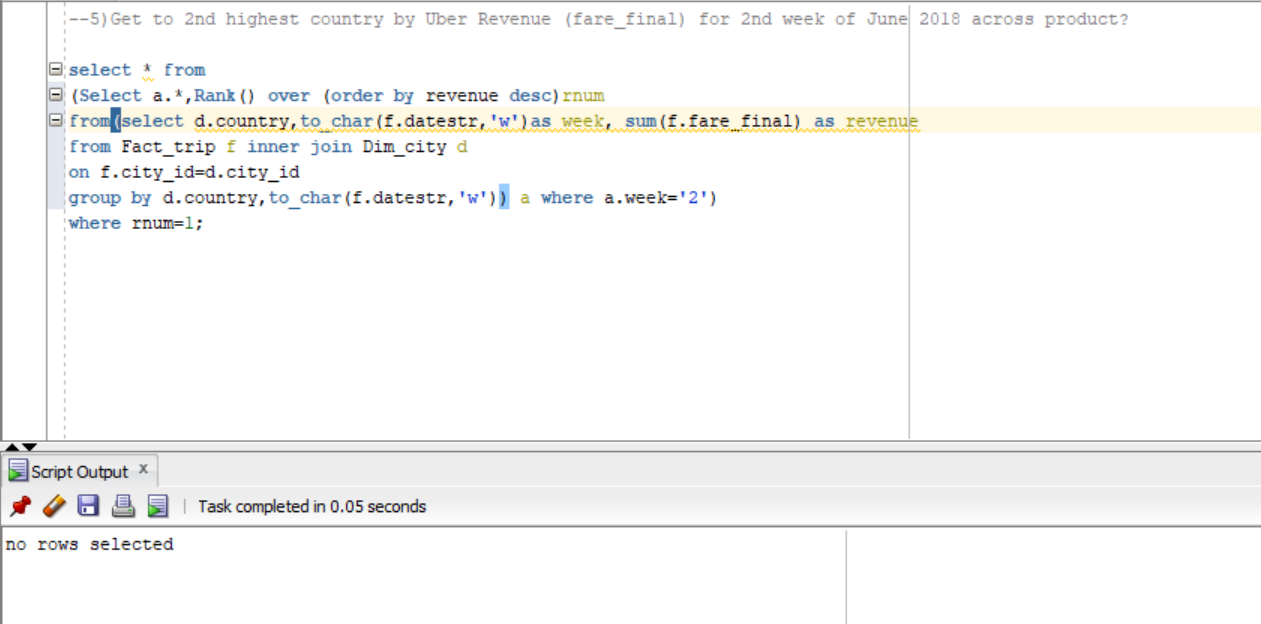




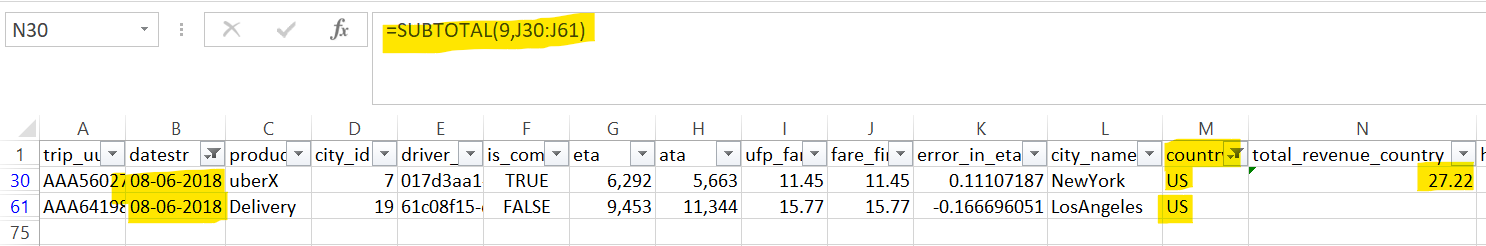




1. Get to 2nd highest country by Uber Revenue (fare\_final) for 2nd week of June 2018 across product?



By analyzing Excel sheet we can see given snip as answer where US has the 2nd highest Uber Revenue (fare\_final) with 27.22



1. Get WOW growth % for US region for June Month. WOW- Week over week. Growth % = ((Current week fare final - previous week fare final) / previous week fare final) \* 100

As we were given with 1week 1day data it’s not possible to find WOW( week-over-week) growth %

1. Any anomalies you identified in the provided dataset and a brief description of how you identified them and why do you think they are anomalies?

I have found single anomaly in Fact\_trip dataset as range of given dates for the service is not sufficient for calculating WOW growth % by this anomaly we can’t able to find the answers for many queries as it’s been showing as no rows are selected.

Kind regards,

Durga Prasad Choudhury