Time DotCom Data Engineering task

# Context:

Kaggle has many datasets available online, we will use one of these datasets to demonstrate your data engineering skills, There are no business rules & guidelines other than to show us what you’re truly made of. It can be as simple or as complex as you want it to be. Just ensure **it's 100% your own original work, we will check and validate**.

The whole point to show us your strengths, try and highlight then using this exercise.

Please use a version control so we can see the history of how you went about it, rather than just uploading the complete project to Github/Bitbucket.

# Dataset:

**Chicago Taxi Trips**

Taxicabs in Chicago, Illinois, are operated by private companies and licensed by the city. There are about seven thousand licensed cabs operating within the city limits. Licenses are obtained through the purchase or lease of a taxi medallion which is then affixed to the top right hood of the car. Source: <https://en.wikipedia.org/wiki/Taxicabs_of_the_United_States#Chicago>

**Content**

This dataset includes taxi trips from 2013 to the present, reported to the City of Chicago in its role as a regulatory agency. To protect privacy but allow for aggregate analyses, the Taxi ID is consistent for any given taxi medallion number but does not show the number, Census Tracts are suppressed in some cases, and times are rounded to the nearest 15 minutes. Due to the data reporting process, not all trips are reported but the City believes that most are. See <http://digital.cityofchicago.org/index.php/chicago-taxi-data-released> for more information about this dataset and how it was created.

This is the link for the dataset:

https://www.kaggle.com/datasets/chicago/chicago-taxi-trips-bq

# The Task:

The dataset is already available on BigQuery, however we want to see your ingestion skills, so you can export the dataset to a google bucket first before you start working on it

We are expecting a full automation of the process, when it comes to ingesting, transformation and loading so show us some Airflow skills, show us as well some DBT / GCP DataForm for transformation and data modelling as well.

We want to know the following

1. Who are the top 100 “tip earners”, the taxi IDs that earn more money than the others for the last 3 months.
2. Who are the top 100 “overworkers”, taxi IDs that work more hours than others without taking at least 8 hours break and regularly have a long shift.
3. Bonus Question: do you think the public holidays in US had an impact on the increase/decrease the trips

You can use looker studio for visualizing your findings, also from the dataset what other insights you found that can be helpful for the company.

GCP gives you free 300 USD credits so please use one to deploy your code. Please make sure you use CI/CD for deploying your code