# Onboarding Test:

This assessment is to determine the likelihood of a candidate being successfully on-boarded as part of the Hyve Data team.

The data relating to these questions is attached, you should have 5 files (after unzipping):

* Readme.txt
* Cell\_tower.csv
* Subscribers.csv
* Transactions.csv (Limited to the 2 and a half months of 2019)
* Subscription\_Data.csv

The assessment is split into three sections weighted as follows:

1. Analysis 40%
2. Visualisation 40%
3. Scripting 20%

Depending on the skills and background of the candidate the weights can be shifted.

The results can be returned in PDF, but source code must be available on request and candidate should explain/ state any assumptions taken with the data. Candidates that can use Tableau and Python for Section B and Sections C & D, respectively, should please do so but are free to use any tools of their choice.

Section A:

1. What is the best performing service per Provider? (Revenue, Subscribers)
2. Which day has the most Active Subscribers per Provider? (Active=period between subscription\_start and subscription\_end)
3. Do any users have subscription across different Providers and or Services (Please give the corresponding user\_ids)?

Section B

1. Create graphs/dashboards that visualise Provider performance per week (Revenue, subscribers)?
2. Create graphs/dashboards that that show subscription acquisition per week (a subscription is acquired on the subscription\_start date), subscription churn (a subscription churn on the date of the subscription\_end) and net subscribers per week?
3. Please plot Active subs per week per Provider for comparison purposes.

Section C

1. Write a script that uses a MAPs API to get country and City for each user based on the input coordinates from the Cell\_tower.csv. The output should be a csv with the following columns user\_id, lat, lng, city, country
2. Which City has the most Active subscribers per product?

Section D

1. Develop a subscription quality scoring methodology and apply it to the data in Subscription\_Data.csv. Please provide your methodology (detailed description and code) as well as your output/results file.

*Note: Data date range is 2020-07-16 to 2021-05-02.*