# TASK BRIEF

**Sub-Task 1:**

Perform some exploratory data analysis. Look into the data types, data statistics, specific parameters, and variable distributions. This first subtask is for you to gain a holistic understanding of the dataset. You should spend around 1 hour on this.

**Sub-Task 2:**

Verify the hypothesis of price sensitivity being to some extent correlated with churn. It is up to you to define price sensitivity and calculate it. You should spend around 30 minutes on this.

**Sub-Task 3:**

Prepare a half-page summary or slide of key findings and add some suggestions for data augmentation – which other sources of data should the client provide you with and which open source datasets might be useful? You should spend 10-15 minutes on this.

For your final deliverable, please submit your analysis (in the form of a jupyter notebook, code script or PDF) as well as your half-page summary document.

Note: Use the 2 datasets within the additional resources for this task and if you're unsure on where to start with visualising data, use the accompanying links. Be sure to also use the data description document to understand what the columns represent. The task description document outlines the higher-level motivation of the project. Finally, use the eda_starter.ipynb file to get started with some helper functions and methods.

**If you are stuck:** Think about ways you can define price sensitivity. Make sure to think of all possible ways and investigate them.

**Estimated time for task completion: 1.75 hours depending on your learning style.**

**Making graphs in Python:**

**1. Matplotlib : Visualization and Python –** https://matplotlib.org/

**2. Seaborn : Statistical Data Visualization –** https://seaborn.pydata.org/index.html

**Making graphs in R:**

**1. Ggplot 2–** https://ggplot2.tidyverse.org/

**2. Plotly R Open Source Graphing Library –** https://plotly.com/r/