V Mobile Data Use Case

V Mobile, as a mobile operator runs a campaign for free minutes for every 60 minutes, per subscriber, each week. Marketing has asked for a weekly report that will contain a list of subscribers, total revenue, total call count for the subscribers that qualified to get the free minutes each week.

The fundamental problem is that marketing is not aware of how many subscribers qualified in the previous week so that they can monitor the campaign performance. In order to qualify, a subscriber must generate a revenue of R30 or more for that week. This revenue can either be from SMS or normal voice calls.

You have the advantage that all data coming from the source systems are clean and accurate, so you do not need to do data cleansing as part of your project. The data provided by the source systems is complete, so you have everything you need to provide the report that Marketing needs.

The Mediation source system that contains SMS and Voice Calls will send you an extract every Monday morning at 04:00 am. The data set will contain raw level data for the past week. [Monday 00:00:00 to Sunday 23:59:59]

Mediation is a process that converts call data to pre-defined layouts that can be imported by a specific billing system. Part of this process is assigning the correct charge to each call transaction and makes it possible to distinguish between free calls and revenue generating calls. In the data provided as part of this project, the 60 minutes free calls that were done by the subscribers that qualified have been excluded from the data set provided by the source system.

Business is also concerned that there are multiple source systems that contain client information within the organisation after the recent amalgamation. The same client can exist in multiple source systems depending on the original mobile company they subscribed with.

Of the various client source systems, V Mobile must be seen as the master, and most trusted source record. All clients will keep only one instance of a cell phone number if there are duplicate numbers found during the client consolidation process.

Then use the data and build the following:

1. Write a Python/SQL script that combines all subscriber information from the three difference subscriber systems.
2. The data must be stored in a **table** that contains the below attributes.
    - First Name
    - Last Name
    - Cell Phone Number
    - Date of Birth
    - Region
    - SIM Activation Date
    - Source System Name [“Name of the system that provided this record”]
    - Is Master Record [“Flag indicating the record that is known as the master record due to business rules see”]

- Master record business rules:
  - The most trusted source of data is V Mobile. If a subscriber exists in VMobile make that record the master record for that subscriber.
  - If the subscriber does not exist in VMobile but exists in BlueMobile or ArrowMobile, you need to compare the dates of SIM Activation Date and mark the master record as the record that has the max/latest activation date.
  - For ties, BlueMobile takes priority over ArrowMobile.

Your solution must produce the following outputs:

1. A weekly report showing a list of subscribers, the total revenue generated by each user, how many SMS's each subscriber sent and how many voice calls they made. Indicate the reporting date and week start and end dates using the *yyyymmdd* date format. The audience for the report is Marketing to support their weekly operations.
2. A visualisation to interactively analyse trends in the data, covering the following analytical questions:
   a What trends are there in qualified subscriber numbers over time? Remember, a subscriber can only qualify once a week.
   b Which city/area has the most qualifying subscribers each week? The challenge is that the call and SMS records that come from the source system have the location ID but does not have the actual location name. The location name resides on another list.

The final presentation to business:

1. A formal presentation is required at the end of your project. The presentation will need to include the scope and process you followed for this project, as well as the conclusions and recommendations you came to after completing the project.

2. A recording of you presenting your findings must be submitted as part of the project artifacts.

The following technology is permitted to be used for each stage of the project.

	Data Analysis	Data Preparation	Data Visualisation	Client Presentation
SQL Server	X	X		
Python	X	X		
Power BI	X		X	
Power Query	X		X	
Excel	X		X	
PowerPoint				X
Canva				X

All the project information/evidence/artifacts must be stored on a private GitHub as per the attached Project Documentation provided.