**Data Engineer Use Case**

Companies lose billions of dollars every year to credit card fraud. Credit card data can be stolen by criminals using a variety of methods. For instance, bluetooth enabled data skimming devices can be placed on the card reader. The data might be stolen in a mass breach by hackers of a large retailer.

A leading financial institution needs to develop a data intensive application that contains financial transactions. Some of these transactions have been reported as fraudulent (reported as 1 under ‘Fraud Class’ in the attached dataset) and others as non-fraudulent (reported as 0 under ‘Fraud Class’ in the dataset). Below is the data dictionary for the 6 month US based dataset (attached credit\_cards\_frauds1.csv) –

* DateTime: Timestamp when the fraud was detected
* Amount: Transaction Amount
* Fraud Class: 1 for fraudulent, 0 otherwise
* Due to confidentiality issues Variable A-E are transformed by using a data transformation technique
* State: The US state where the fraud activity was recorded
* Type of credit card: Master Card/Visa / Discovery service providers
* Device: The device type used when the fraud activity was recorded

Please find below the list of tasks you need to perform with the attached dataset-

1. Upload the dataset in SQL and **paste a snapshot of the schema** (as it appears in your SQL platform) for this below.

USE [Northwind]

GO

/\*\*\*\*\*\* Object: Table [dbo].[credit\_card\_frauds1] Script Date: 04-01-2019 14:11:23 \*\*\*\*\*\*/

SET ANSI\_NULLS ON

GO

SET QUOTED\_IDENTIFIER ON

GO

CREATE TABLE [dbo].[credit\_card\_frauds1](

[Variable A] [float] NULL,

[Variable B] [float] NULL,

[Variable C] [float] NULL,

[Variable D] [float] NULL,

[Variable E] [float] NULL,

[Amount] [float] NULL,

[State] [varchar](50) NULL,

[Type of credit card] [varchar](50) NULL,

[Device] [varchar](50) NULL,

[Date time] [datetime] NULL,

[Fraud Class] [int] NULL

) ON [PRIMARY]

GO

1. Write a SQL query to extract the following information:
   1. National and state-wise number of frauds (**please paste the SQL query developed below**)

National

select COUNT([Fraud Class]) as Frauds

FROM [Northwind].[dbo].[credit\_card\_frauds1]

where [Fraud Class]=1

State-wise

select [State], COUNT([Fraud Class]) as Frauds

FROM [Northwind].[dbo].[credit\_card\_frauds1]

where [Fraud Class]=1

group by [State]

* + 1. Identify the state with the maximum number of frauds

select a.[State] from

(

select [State], COUNT([Fraud Class]) as Frauds

FROM [Northwind].[dbo].[credit\_card\_frauds1]

where [Fraud Class]=1

group by [State]

)

a

where a.Frauds in (

select MAX(b.Frauds) from

(

select [State], COUNT([Fraud Class]) as Frauds

FROM [Northwind].[dbo].[credit\_card\_frauds1]

where [Fraud Class]=1

group by [State]

)

b

)

* + - 1. For the identified state above please draw insights around fraud activities observed (descriptive analytics). To do this task please follow the following process-
         1. Write a SQL query to extract the fraud data for this state (**please paste the SQL query developed below**) and export it to an Excel file (**please attach this Excel file in your reply**)

select \* from [Northwind].[dbo].[credit\_card\_frauds1]

where [State] = 'California' and [Fraud Class] = 1

* + - * 1. Use R/Python to read the Excel file from part a. above and analyse this data to report insights derived from this data (you will need to create the plots in your tool of choice)

1. **Prepare a PowerPoint deck to report your data analysis and insights derived** from the above (use charts/plots at every place possible- make sure they are labelled appropriately so that they are self-explanatory). For this PowerPoint presentation make sure you prepare a proper storyline such that you can report meaningful and actionable insights to your client. (attached is the PowerPoint template for your convenience)
2. **Finally write an email to your client with the highlights of your findings/insights from this data**
3. Please note the expected documents in your response to this assignment are-
   1. The current word document with all the SQL queries pasted, wherever required
   2. The R/Python file used for data analysis
   3. The Excel extract requested
   4. The PowerPoint presentation with your analysis
   5. Client email