Computer Science Department

Capstone Project

***Euro-Financial Data Distribution***

CSC 521

Dec 2021

By

*SAMUEL KAWUMA*

CSC 521 Project Supervisor

*FATEMA NAFA*

**Contents**

1. Student Objectives x
2. Problem Specification x
3. Solution Process and Design
4. Requirement Collection and Analysis

1. Data Sequence Diagram x

1. High level Architectural Diagram x

1. UI Mockups x

1. Benchmark Specification x

1. Tool List x

1. Time Schedule x

1. Grading Scheme for CSC521 x

1. Deliverables x
2. References x

# Student Objectives

The main objective of my capstone course is to enable me to integrate the knowledge I have gained as a result of pursuing my computer science degree program in college. I plan to integrate the skills and concepts learned systematically during my semester

As a result, I should be able to integrate all aspects of the course which includes the theory, practical skills, and communication skills. I plan to combine the diverse skills acquired in class and apply them work environment. The capstone project thus should act as a self-assessment mechanism that reminds me of the expected learning outcomes.

My Capstone project should, therefore, focus on helping me to perfect technical skills in handling a given task. This requires me to have hands-on experience in the fields of specialization that I have chosen to undertake

**Problem Specification**

My project as an aspect of Data Science uses Visualization to display financial data of individuals in certain European countries. The data sets would include Customer Id, CreditScore, Geography/ Country,Age, Account Balance, Has Credit Score,Estimated Salary.

My main goal in this project is to make it easier to identify patterns, trends and outliers in my data set.

I would best present my project with information graphics information visualization and statistical graphics.

# Solution Processes and Design

1. **Requirements and Analysis**

Stake holders

The two types of stakeholders were. Identified as customers and business entities

**Customers**: This can range from everyday user like students’ corporate workers (. Working class). These would use the application to typically connect financial records such as csv where they find and select appropriate data that can be used in order to create a visualization that answers a particular research question.

**Business entities:** Which allows allow stakeholders with little or no knowledge of data science, work independently on data, and derive some findings that might assist their day-to-day business decisions

The input to the application is a .csv file served as the raw data for the application. The raw data consists of many attributes. Here is a screenshot of the titles of the .csv file.

**Customer Id**

contains random values that are independent and unique has no impact.

**Surname**

the surname of a customer is independent value and hence has no impact

Using Salary data, we can visualize and relate age, sex and credit score with customer estimated salary. We can draw out an analysis that probably a higher age tends to correlate with higher salary or vice versa. We can derive that certain sex has a higher salary than the other sex. (for example males have higher estimated salary than females… Also, you can analyze that having higher salary can lead to higher credit score or having lower Salary can lead to one having lower credit score.

**Gender**

Gender can have a correlation with credit score you can assert that either males tend to have a higher credit score than females or vice versa. Also, you can visualize sex against customer balance and from this analysis you can visualize what gender tends to have higher account balance than the other.

**CreditScore**

Using credit score data we can relate credit score to first Age, then sex and geographical area.

We can visualize what sex / gender has a higher credit score. We can also visualize what Age bracket tends to have a higher credit score.

**Geography**

Using Geography we able to relate customer balance and their regions/ Countries they come, in this case using visualization you can determine what regions(countries) tend to have higher

**EstimatedSalary**

**HasCrCard**

Using credit card data, we can visualize and see that certain age group tend to have credit cards and the other seems not to have.

**Visualization Outputs**

The application will deliver its outputs as the follows

**Credit score Distribution by Age**

Here a system user will be able to display a bar chart that reveals average credit score against age range. Using this data, a user can tell what age group has a higher credit score and that with a lower credit score.

**Estimated Acc. Balance Distribution by Gender**

Here a system user will be able to display a pie chart that shows estimated acc. balance by gender. With this data a user can know what gender appears to have a higher account balance than the other.

**Estimated Acc. Balance Distribution by Age**

Here a system user will be able to display a histogram that shows estimated acc. balance by Age. With this data a user can know what age group appears to have a higher account balance than the other.

**Credit score Distribution by country**

Here a system user will be able to display a bar chart that reveals average credit score against Countries. Using this data a user is able to tell what country has people that tend to have a higher credit score and that with a lower credit score.

**Salary Distribution by Age**

Here a system user will be able to display a histogram that shows Salary by Age. With this data a user can know what age group appears to have a higher salary than the other.

**Credit Card Distribution by Gender**

Here a system user will be able to display a pie chart shows credit card distribution by gender. With this data a user can know what gender appears to have a higher credit card possession than the other.