CSC8643: Data Management and Exploratory Data Analysis

by Matthew Taylor

20th November, 2023

Contents

1 Abstract **4**

2 Introduction **4**

3 Business Understanding **4**

3.1 Business Objective 4

3.2 Access Situation 4

3.2.1 Inventory of Resources 4

3.3 Determine Data Mining Goals 4

3.4 Produce Project Plan 4

4 Data Understanding **6**

4.1 Collect Inital Data 6

4.2 Decribe Data 6

4.3 Explore Data 6

4.4 Verify Data Quality 6

5 Data Preparation **6**

5.1 Select Data 6

5.2 Clean Data 6

5.3 Construct Data 6

5.4 Intergrate Data 6

5.5 Format Data 6

6 Modeling **6**

6.1 Select Modeling Techniques 6

6.2 Generate Test Design 6

6.3 Build Model 6

6.4 Assess Model 6

7 Evaluation **6**

7.1 Evaluate Results 6

7.2 Review Proccess 6

7.3 Deterime next steps 6

8 Deployment **6**

8.1 Plan Deployment 6

8.2 Plan Monitoringand Maintenance 6

8.3 Produce Final Report 6

8.4 Review Project 6

9 Analysis **8**

1 Abstract

2 Introduction

3 Business Understanding

3.1 Business Objective

Newcastle University designed a massive open online course entitled "Cyber Security: Safety At Home, Online, and in Life". The Primary objective of this course was to provide a free resource that was easily accessible via the online provider FutureLearn on the topic of cyber security, and how the learners can protect their digital data.

The course was ran 7 times, with the majority of users signing up at the beginning of the the first and second academic semesters from 2016 through 2018. The tasks of the course were exclusively online, and could be completed asynchronously.

The FutrureLern system recorded a vast about of data, covering a range of variables, including demographics, and engagement with the course, and because each user had a unique identifier, and their demographic and engagement data was recorded against their ID, there is a wide range of analysis that can be done.

3.2 Access Situation

3.2.1 Inventory of Resources

This project is being undertaken with a limited number of resources. There is a vast amount of data including demographics, and engagement with the course, as well as survey responses. Having an extensive amount of raw data is helpful to the aims, but requires additional resources to handle. For this project, I am a team of one, with no data science or modelling expertise.

3.3 Determine Data Mining Goals

The specific goal of this analysis is to dig into the country demographics of the learners, and determine if there is a pattern that could lead to increasing the number of people who take the course, or increase the engagement of current users.

3.4 Produce Project Plan

Due to my limited experience in data modeling, I will deploy some limited analysis tequiques.

4 Data Understanding

4.1 Collect Inital Data

4.2 Decribe Data

4.3 Explore Data

4.4 Verify Data Quality

5 Data Preparation

5.1 Select Data

5.2 Clean Data

5.3 Construct Data

5.4 Intergrate Data

5.5 Format Data

6 Modeling

6.1 Select Modeling Techniques

6.2 Generate Test Design

6.3 Build Model

6.4 Assess Model

7 Evaluation

7.1 Evaluate Results

7.2 Review Proccess

7.3 Deterime next steps

8 Deployment

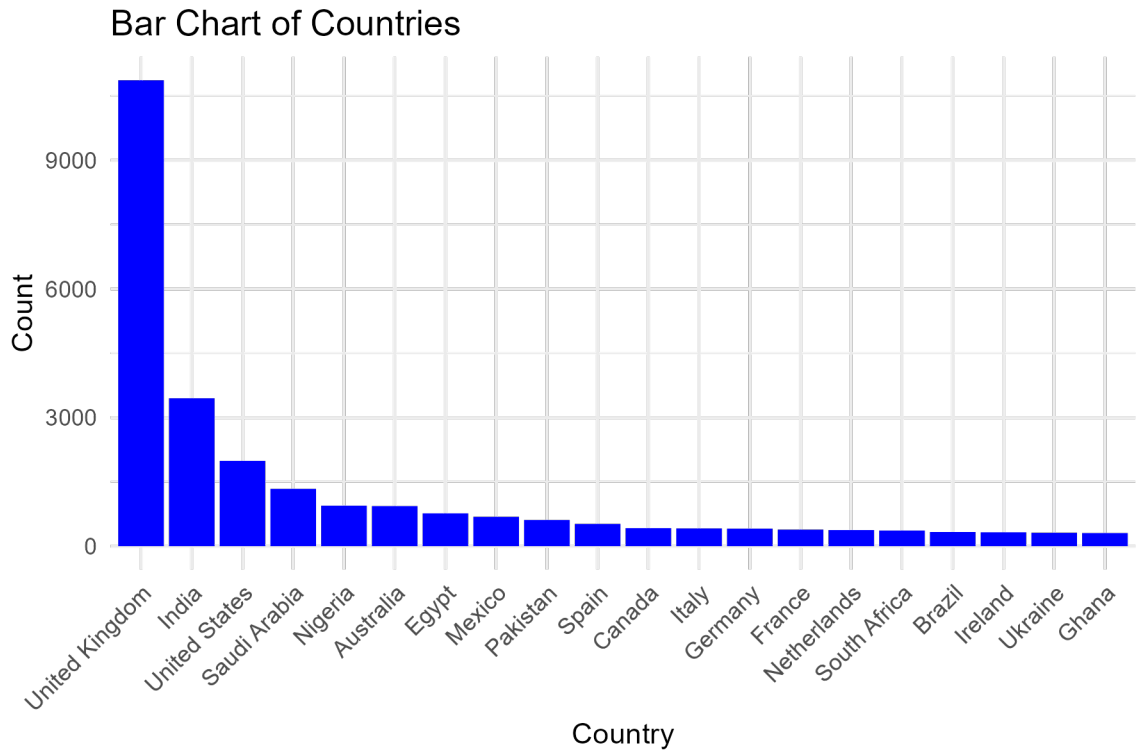
8.1 Plan Deployment

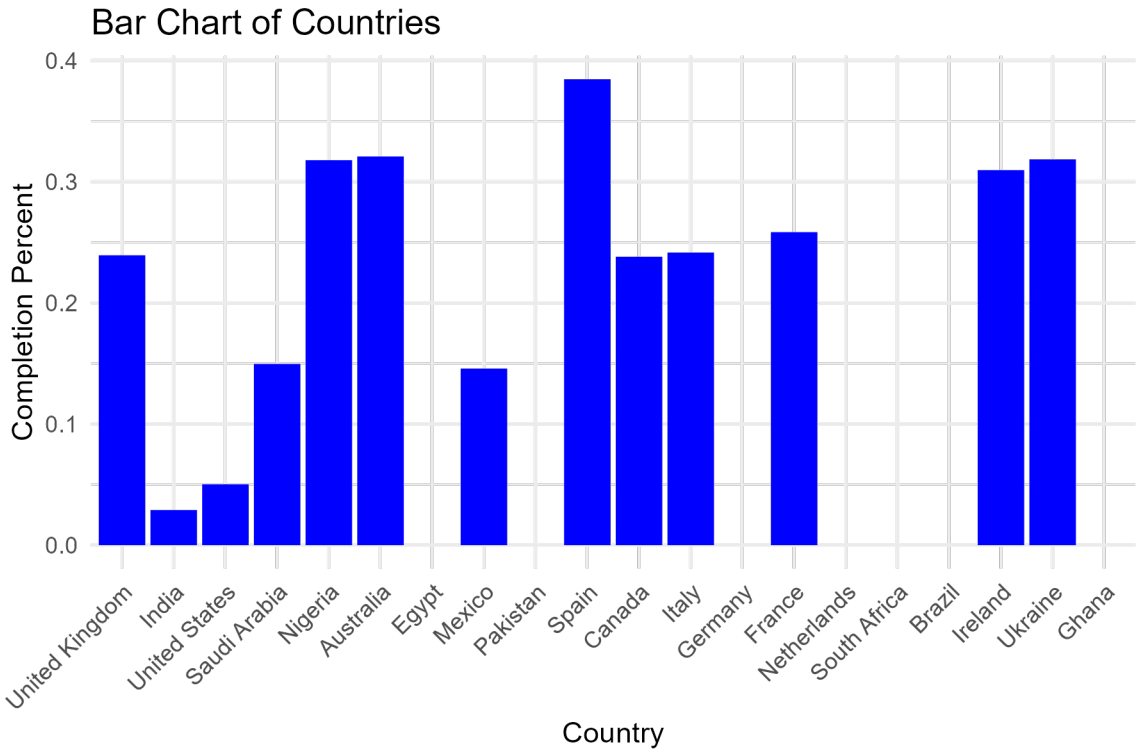
8.2 Plan Monitoringand Maintenance

8.3 Produce Final Report

8.4 Review Project

here is a bar chart:





9 Analysis

talking talking talking...