Assignment01

Fei Yun

8680643

**Short Answer**

1. How many data records are in your final dataset?

185

2. How many data records did you manually remove during the cleansing process?

19

3. How many countries of origin are contained in your final dataset? List them.

18 countries

{'present-day France', 'Netherlands', 'France', 'Costa Rica', 'United States', 'present-day Uzbekistan', 'Mexico', 'Papua New Guinea', 'Iran', 'India', 'Egypt', 'Indonesia'}

**BI and Data related Questions**

1. What are the 3 Vs of Data? [1 marks]

Volume,Velocity,Variety

2. List capabilities of Business Intelligence systems. [5 marks]

Data collection, data visualizations, reporting, data analysis, data ingestion

3. Different types of data with example for each type. [3 marks]

Structured: database

Unstructured: videos,pictures

Semi-structured: Emails

4. Define data visualization. [2 marks]

Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data.

5. What is BI system? [1 marks]

Business Intelligence utilizes analytics, data mining, visualization, tools and infrastructure to help organizations to make more data-driven decisions

6. What are the 5 C's of Data for data preparation and the purpose of each? [5 marks]

Clean:

Removeoreditincomplete,incorrect,inaccurateor irrelevant datapoints

Consistent: Therelationshipsbetweendatapointsandhowtocomparethemmustbemadestandardthroughout thesystem

Conformed:

Theprotocolsforhowto analyzethedatamustbe madestandardthroughoutthesystem

Current:

Determinehowquicklynewdatawillbemade available

Comprehensive:

Ensurethatthedatarequiredto supportandecision is available