**Data Warehousing and Data Mining**

**Assignment1**

**M.Naveen Kumar**

**19BCS107**

Active Learning – Understanding Data Mining

1.What is a purpose of data mining?

Data mining is the process of finding anomalies, patterns and correlations within large data sets to predict outcomes. Using a broad range of techniques, you can use this information to increase cut costs, improve customer relationships, reduce risks

2. Compare Gold mining, Coal mining and datamining?

Gold mining is a global business with operations on every continent, except Antarctica, and gold is extracted from mines of widely varying types and scale. Coal mining, extraction of coal deposits from the surface of Eanh and from underground Data mining is the process of finding anomalies, patterns and correlations within large datasets to predict outcomes

3.1llustrate various steps involved in datamining?

1. Data Cleaning.

2) Data Integration.

3) Data Reduction.

4) Data Transformation.

5) Data Mining.

6) Pattern Evaluation.

7) Knowledge Representation

4.List Any four datamining tools ?

* Oracle Data Mining.
* IBM SPSS Modeler.
* Knime.
* Python.
* Orange.
* Kaggle.
* Rattle.

5.Data Mining is a Multidisciplinary Field? Justify Your Answer?

Data mining is a multidisciplinary field drawing works from statistics, database technology, artificial intelligence, pattern recognition, machine learning, information theory, knowledge acquisition, information retrieval, high-performance computing, and data visualization.

**Active Learning -Understanding Dataset**

1.Difference between data, dataset, database?

Data are plain facts. The word "data" is plural for "datum." When data are processed, organized, structured or presented in a given context so as to make them useful, they are called Information. It is not enough to have data .Dataset vs Database will provide us with the differences by comparing both in the professional field. Dataset is a structured collection of data that is generally associated with a unique work body. And the database is an organized collection of data stored into multiple datasets, and these datasets are stored and accessed electronically from a computer system which allows being easily accessible, for manipulation and updating. These two terms are used loosely and have different definitions overall. Database tends to manage the collection of statements whereas a dataset is a fixed collection of propositions. Here, we shall compare the dataset and database, listing down the similarities and differences. Also, will get through the key differences between the dataset and database.

2.Different Types of Attributes in dataset?

Here are the different types of Attributes

1. **Qualitative Attributes** such as Nominal, Ordinal, and Binary Attributes.
2. **Quantitative Attributes** such as Discrete and Continuous Attributes.

3.List Attribute type for each attribute separately and justify?

1. Employee\_id - Normal Attribute
2. Department - Normal Attribute
3. Gender - Normal Attribute
4. Designation - Normal Attribute
5. Employment

Status - Ordinal Attribute

1. Basic pay - Ordinal Attribute
2. Height - Ordinal Attribute
3. Weight - Ratio Scaled Attribute
4. Age - Ratio Scaled Attribute

10.Telephone no - Numeric Attribute

11.Experience in Year – Numeric Attribute

12.Permanent since – Numeric Attribute

13.Employee Rating – Numeric Attribute

14.Employee Name – Normal Attribute

15.Joining Date – Ordinal Attribute

4.Select the Attributes that is relavant for Offering Promotion to Employee

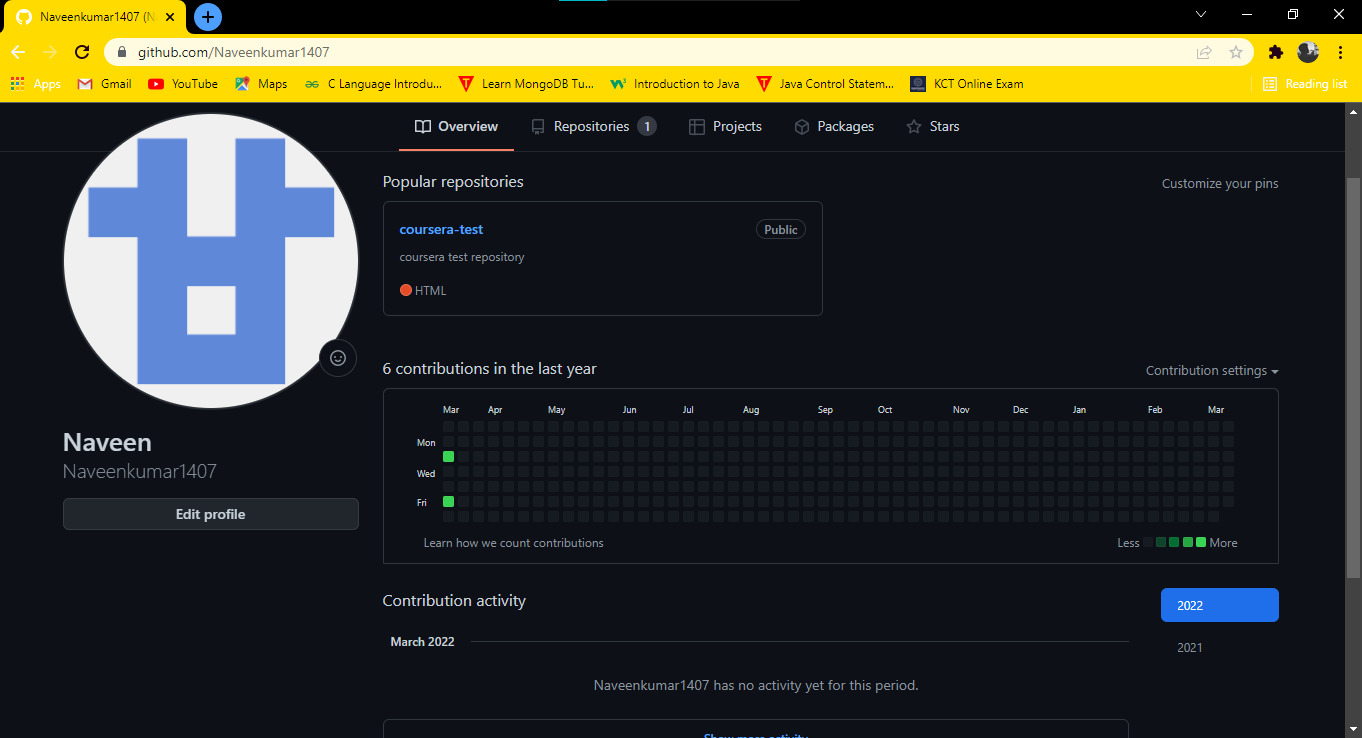
Justify Answer?

* Employee\_id
* Designation
* Employee status
* Experience in Years
* Basic pay
* Permanent since
* Employee Rating

**Individual Task 3**

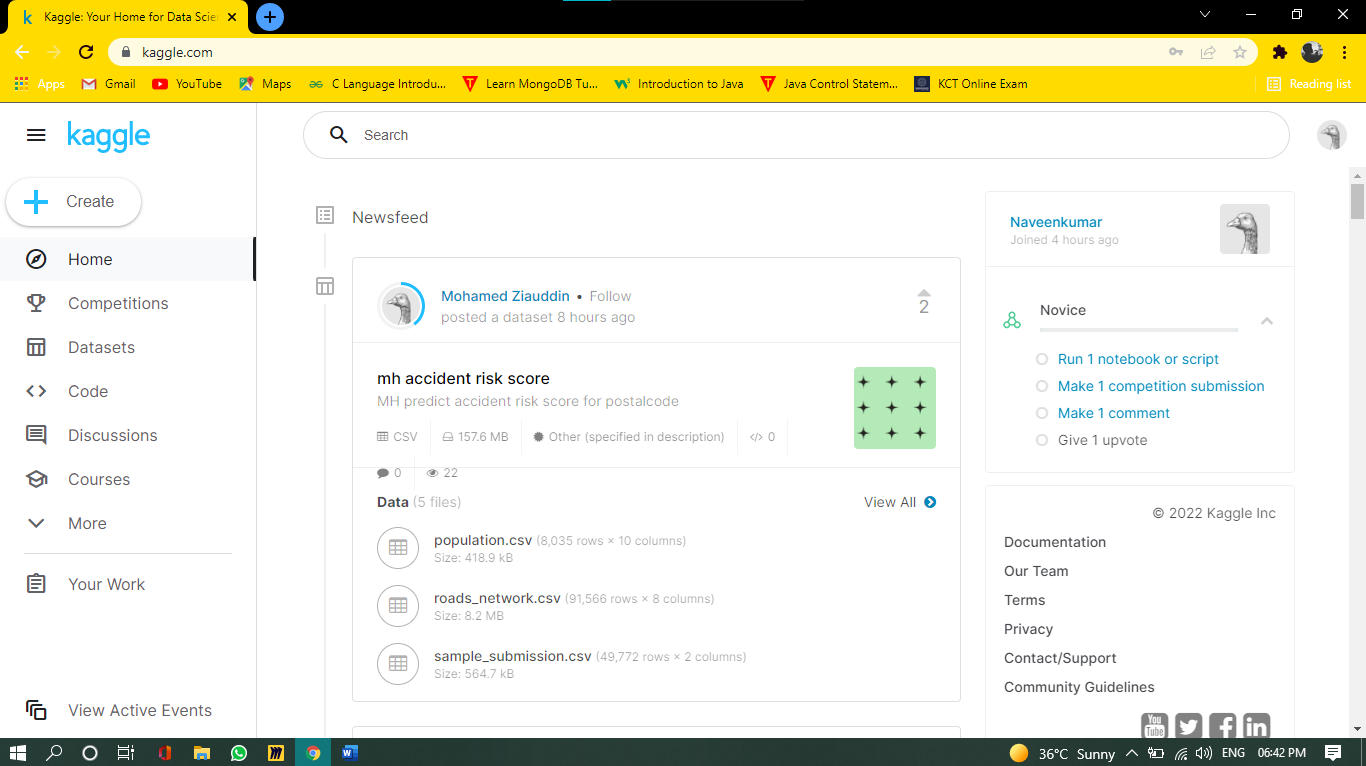
1.Create a Github account and specify id

Id: <https://github.com/Naveenkumar1407>



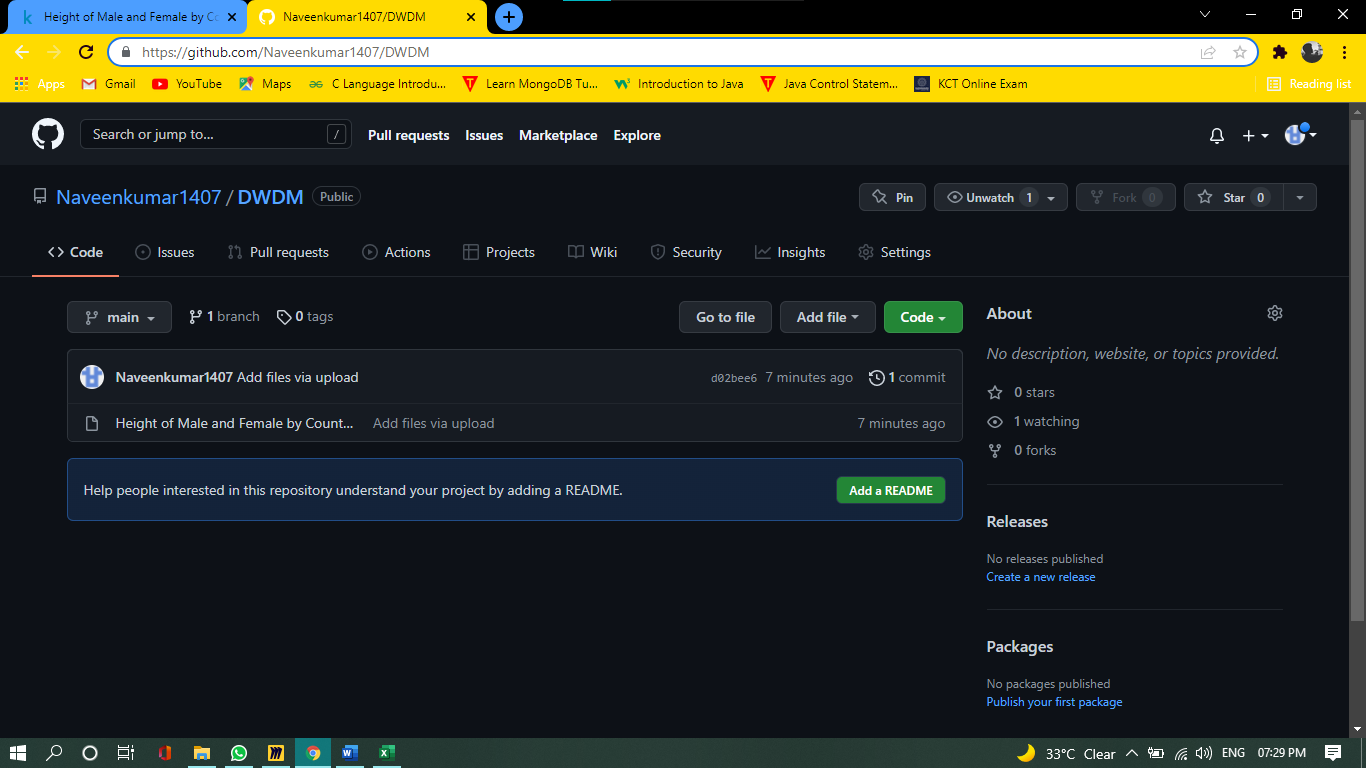
2.Create a Kaggle account and specific id

Id :Naveenkumar



3.Download a dataset from Kaggle and upload in github and provide the link

<https://github.com/Naveenkumar1407/DWDM>



4.In the dataset downloaded List the attributes and types

1. Rank-Ordinal Attribute
2. Country Name-Normal Attribute
3. Male Height in Cm-Ordinal Attribute
4. Male Height in Ft-Ordinal Attribute
5. Female Height in Cm-Ordinal Attribute
6. Female Height in Ft-Ordinal Attribute