Batch: B1 Roll No.: 16010420133 Name: Soumen samanta
Experiment No.:1

Aim: Understanding of the Data

Resources needed: Any RDBMS, EXCEL, Data storage tool

Procedure / Approach /Algorithm / Activity Diagram:

1) Download the large dataset for the purpose of exploration and ensure that dataset has variety of attributes; number of attributes must be at least 25.

Dataset: https://www.kaggle.com/gdaley/hkracing?select=races.csv
Title: Horse Race (Runs)

2) Identify the category of each attribute from the dataset which you have created.

Column Name	Data Type
race_id	Interval-Scaled
horse_no	Interval-Scaled
horse_id	Interval-Scaled
result	Interval-Scaled
won	Binary
lengths_behind	Ratio-Scaled
horse_age	Interval-Scaled
horse_country	Nominal
horse_type	Ordinal
horse_rating	Interval-Scaled
declared_weight	Interval-Scaled
actual_weight	Interval-Scaled
draw	Interval-Scaled
position_sec1	Interval-Scaled
position_sec2	Interval-Scaled
position_sec3	Interval-Scaled
position_sec4	Interval-Scaled
behind_sec1	Ratio-Scaled
behind_sec2	Ratio-Scaled
behind_sec3	Ratio-Scaled
behind_sec4	Ratio-Scaled
time1	Interval-Scaled
time2	Interval-Scaled
time3	Interval-Scaled
time4	Interval-Scaled
finish_time	Interval-Scaled
win_odds	Interval-Scaled
place_odds	Interval-Scaled
trainer_id	Interval-Scaled

jockey_id	Interval-Scaled

3) Identify the attributes which can provide any kind of useful information either collectively or as an individual. Also, discuss the about the information provided by the attribute and how it will be computed?

A) Horse_id , lengths_behind , declared_weight and actual_weight attributes are used to keep a track. Position_sec1 , position_sec2 , position_sec3 , position_sec4 , behind_sec1 , behind_sec2 , behind_sec3 , behind_sec4 provide the information of how much the horse is behind the line and with respect to time.

Post Lab Questions:

1. Compare Discrete and Continuous Attributes. Give at least 5 examples of each.
A)

Discrete	**Continuous**
Countable	Measurable
Bar Graph is used for representation	Histrogram is used for representation
Mutually Inclusive	Mutually Exclusive
It can take only distinct or separate values.	It can take any value in some interval
Example : Shoe size Numbers of siblings Cars in a parking lot Days in the month with a temperature measuring above 30 degrees Number of students in a class A list of a baseball team's seasonal wins Number of different vegetables in a crate	Example: The volume of a gas tank in liters Wind speed in miles per hour The height of buildings in meters Length of a rope in inches Temperature (in degrees, on any measurement scale) The time it takes runners to complete a race in minutes The weight of a crate of vegetables in kilograms

Outcomes:

CO1: Summarize the data

Conclusion:

We learnt about the various data attributes and to identify the category of each attribute from the dataset .