DATA WRANGLING

Capstone project – 1

*Title:*

Graduate Admission 2 - Predicting admission from important parameters

Cleaning steps performed:

Inconsistent column names:

Two column names are inconsistent. It has a space ‘LOR ‘ and Chance of Admit ‘

I have removed the space by renaming the columns as below

df.rename(columns = {'Chance of Admit ':'Chance of Admit'}, inplace = True),

df.rename(columns = {'LOR ':'LOR'}, inplace = True)

Missing values:

There are no missing values

This was checked with df.info() method

Outliers:

There are no outliers

This was checked by comparing aggregate mean and max values

Obtained by df.describe()

A screenshot of a cell phone

Description automatically generated

This was also checked by scatter plot

A screenshot of a cell phone

Description automatically generated

Data Story Telling:

**Steps:**

1. Ask the following questions and look for the answers using code and plots:
   1. Can you count something interesting?
   2. Can you find trends (e.g. high, low, increasing, decreasing, anomalies)?
   3. Can you make a bar plot or a histogram?
   4. Can you compare two related quantities?
   5. Can you make a scatterplot?
   6. Can you make a time-series plot?
2. Looking at the plots, what are some insights you can make? Do you see any correlations? Is there a hypothesis you’d like to investigate further? What other questions do the insights lead you to ask?
3. Now that you’ve asked questions, hopefully you’ve found some interesting insights. Is there a narrative or a way of presenting the insights using text and plots that tells a compelling story? What are some other trends/relationships you think will make the story more complete?