DS 520/CS 520 – Lecture 3

DS-520-50: Data Analy: Conc/Tech
2024 Fall
MONMOUTH CAMPUS

T 7:30 PM - 10:20 PM
9/3/2024 - 12/9/2024
Howard Hall, 306 LECTURE

Professor Arup Das
adas@Monmouth.edu

Disclaimer:

- The views expressed are solely those of the presenter and not affiliated with any other party.
- This presentation is free of copyright violations, and external sources have been appropriately credited.
- T**he content within this presentation is legally protected; unauthorized reproduction, including photography, will result in legal action.**
- **This material is not intended for distribution and must remain solely within the confines of this class. Do not distribute slides or assignments to other students**
- **Using cameras to take screenshots or photographs of the slides is strictly prohibited**.

Course Logistics

Date	Week	Class Format/Location/Time	Topics	Readings Required (Due before class)	Assignment/Quiz
September 3,2024	Week_1	On-Premise/Howard Hall, 206 LECTURE/7:30 PM-10:20PM	Python for Machine Learning Refresher		
September 10,2024	Week_2	On-Premise/Howard Hall, 306 LECTURE/7:30 PM-10:20PM	Machine Learning Workflow and Exploratory Data Analysis (EDA)	Book 1 – Chapter 1, Chapter 2 (Pages 22 – Page 29)	
September 17,2024	Week_3	On-Premise/Howard Hall, 306 LECTURE/7:30 PM-10:20PM	Machine Learning Workflow and Exploratory Data Analysis (EDA)	Book 1 – Chapter 2 (Page 32- 63)	**Project 1 Distributed - Due Sep 27,2024**
September 24, 2024	Week_4	On-Premise/Howard Hall, 306 LECTURE/7:30 PM-10:20 PM	Machine Learning for Regression	Book 1 – Chapter 2 (Page 32- 63)	
October 1, 2024	Week_5	On-Premise/Howard Hall, 306 LECTURE/7:30 PM-10:20 PM	Machine Learning for Regression	Book 1 – Chapter 3 (Page 65- 110)	**Project 2 Distributed - Due Oct 12, 2024**
October 6, 2024	Week_6	On-Premise/Howard Hall, 306 LECTURE/7:30 PM-10:20 PM	Machine Learning for Classification	Book 1 – Chapter 4 (Page 113- 145)/Chapter 6	
October 22, 2024	Week_7	Zoom remote/8:00pm – 10:30 pm	Machine Learning for Classification	Book 1 – Chapter 4 (Page 113- 145), /Chapter 6	**Project 3 Distributed – Due Nov 1, 2024**
October 29,2024	Week_8	Zoom remote/8:00pm – 10:30 pm	Quiz 1 (Cover materials from Week 1-7)	Book 1 – Chapter 3 (Pages 88- 92)	**Quiz 1 – Open Book/Open Notes**
November 5, 2024	Week_9	Zoom remote/8:00pm – 10:30 pm	Feature Selection, Model Selection and Tuning	Book 1 – Chapter 4 (Pages 147- 151)	
November 12, 2024	Week_10	Zoom remote/8:00pm – 10:30 pm	Feature Selection, Model Selection, and Tuning	Book 1 – Chapter 4 (Pages 147- 151)	**Project 4 Distributed – Due Nov 22, 2024**
November 19,2024	Week_11	Zoom remote/8:00pm – 10:30 pm	Unsupervised Learning	Professor Lecture Notes	
November 26, 2024	Week_12	Zoom remote/8:00 pm – 10:30 pm	Unsupervised Learning	Professor Lecture Notes	**Project 5 Distributed – Due Dec 6, 2024**
December 3, 2024	Week_13	Zoom remote/8:00 pm – 10:30 pm	AI Certifications Overview or additional topics spill over from preceding weeks	Professor Lecture Notes	**Quiz 2 Distributed**
December 9, 2024 – Last day of class	Week_14	Quiz 2 Due	Quiz 2 (Covers materials from Weeks 9 -12) and Course wrap-up		**Quiz 2 Due Dec 8, 2024 before midnight EST**

Course Logistics

1. OneDrive link for professor notes and assignments/quiz
2. Check your Monmouth email for announcements
3. Check your Monmouth calendar for Zoom links for office hours and remote lectures
4. My contact information: adas@monmouth.edu, Cell # 917-523-7683
4. Office hours (zoom only) – Friday (EST) 7-7:30 pm EST
5. Assignment submission to professoraruprdas@gmail.com (Notation for files: Assignment_1_Name_of_Student), Colab notebooks ipynb file and html file, all presentation in ppt format.
6. Quiz submission to professoraruprdas@gmail.com (Notation for file : Quiz_1_Name_of_Student. doc , Quiz_2_Name_of_Student.doc)

Study Groups

Group 1	Group 2	Group 3	Group 4
Andrew Captano - Group Leader	Kiran Ramjisingh - Group Leader	Shobharani Polasa- Group Leader	Sashank Vaddiparty- Group Leader
Ryan Sonn	Killariben Limbachiya	John Costa	Katia Bravo Bendezu
Alonso Aguilera	Ezzine Ndumnwere	Vibushan Raju Guduri	Amulya Konduru
Noah Ferker			

Lecture_4 Assignments

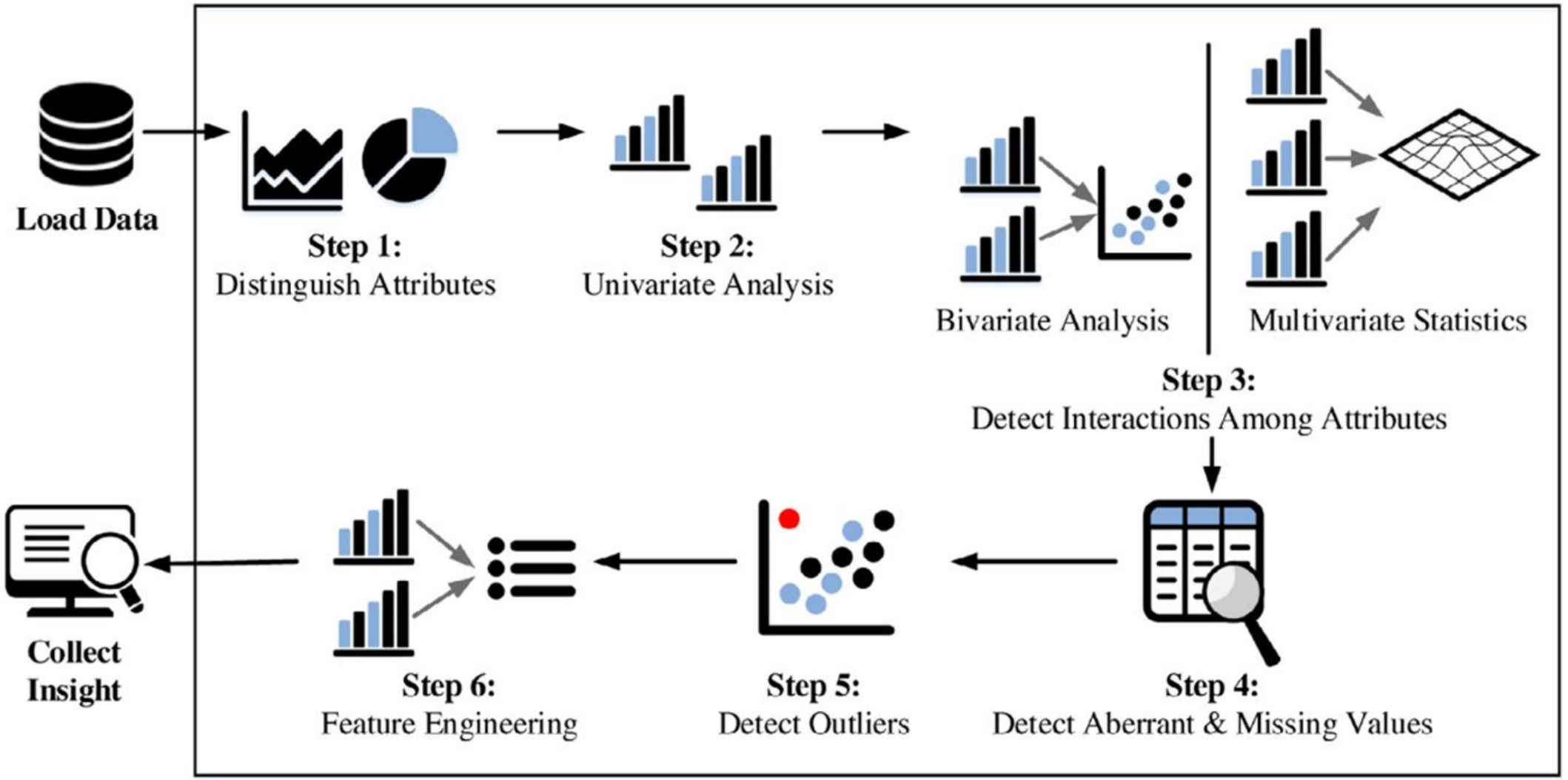
- Kaggle Badges complete
- Review Book 1 – Chapter 2 (Page 32- 63)

Topics Lecture_3

1. Review main topics from Lecture 2
2. EDA in class notebook
3. Assignment_1 Walk_through

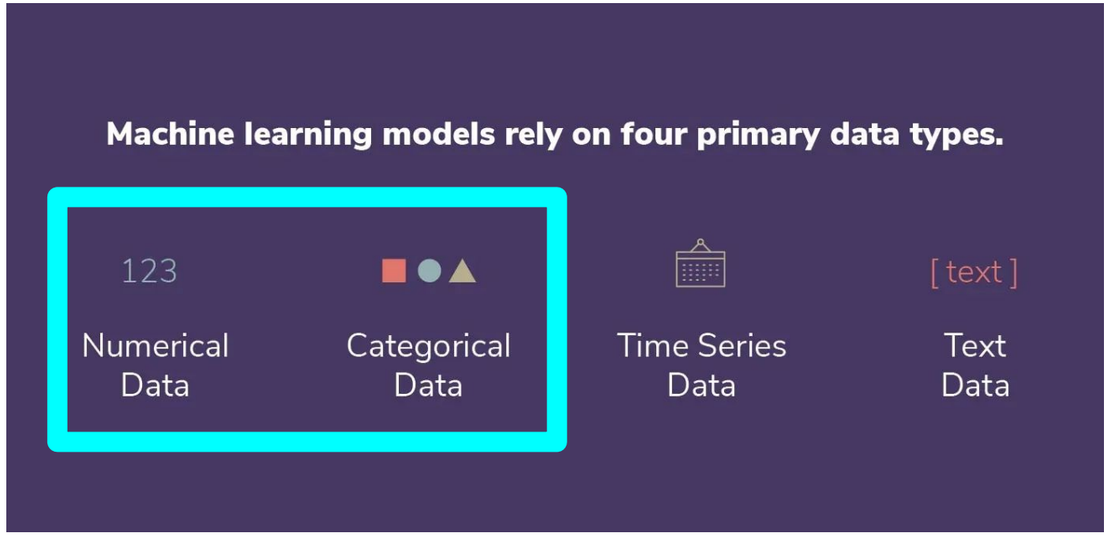
Lecture_2 Refresher

EDA Process



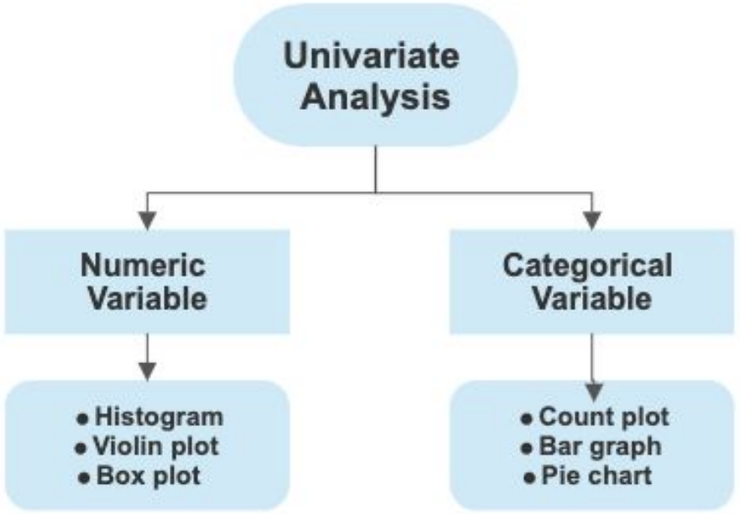
Source: https://devopedia.org/exploratory-data-analysis

Data Types in Python

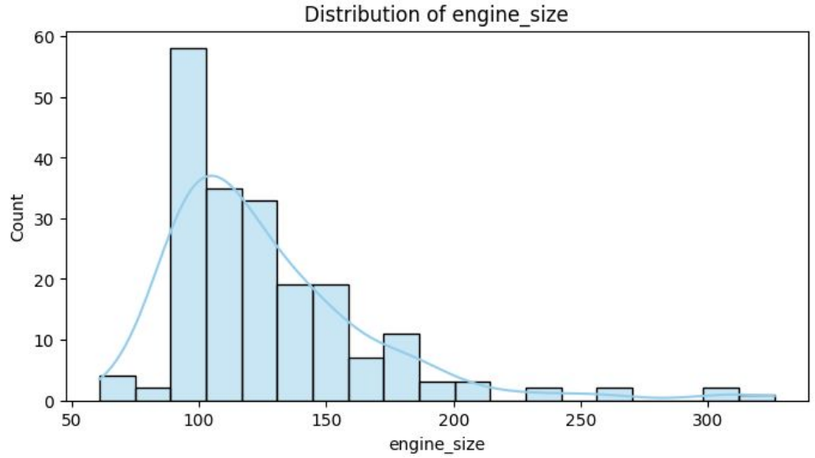


Pandas dtype	Python type	NumPy type	Usage
object	str	string_, unicode_	Text
int64	int	int_, int8, int16, int32, int64, uint8, uint16, uint32, uint64	Integer numbers
float64	float	float_, float16, float32, float64	Floating point numbers
bool	bool	bool_	True/False values
datetime64	NA	datetime64[ns]	Date and time values
timedelta[ns]	NA	NA	Differences between two datetimes
category	NA	NA	Finite list of text values

Univariate Analysis

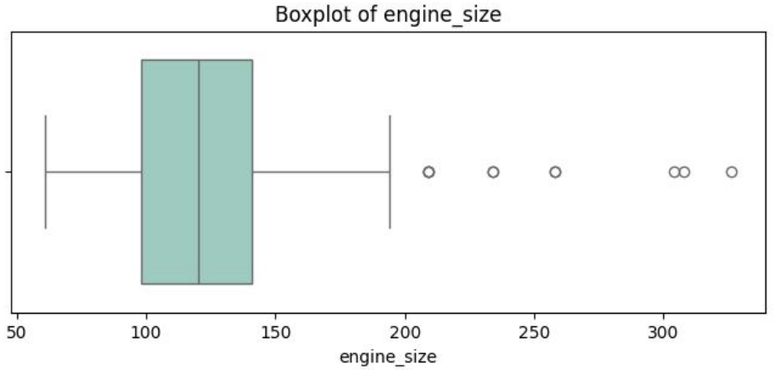


Histogram

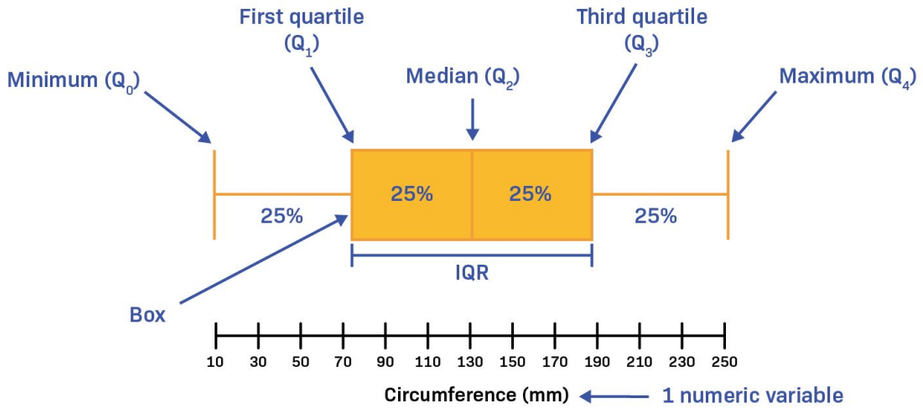


Distribution of variables

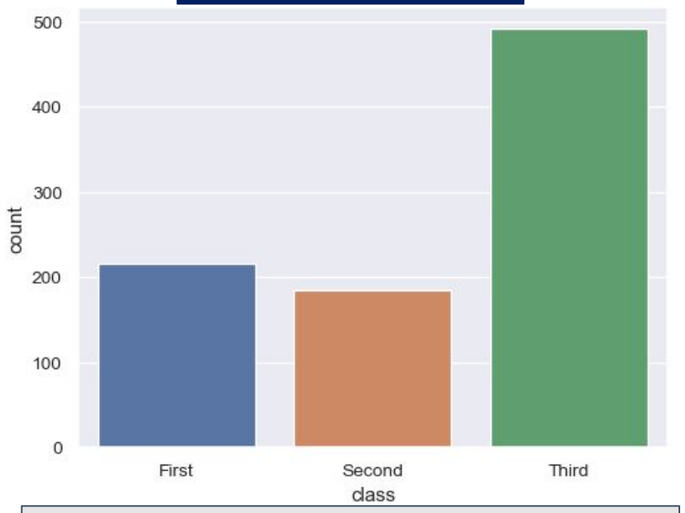
Box Plot



Variable statistics + Outliers

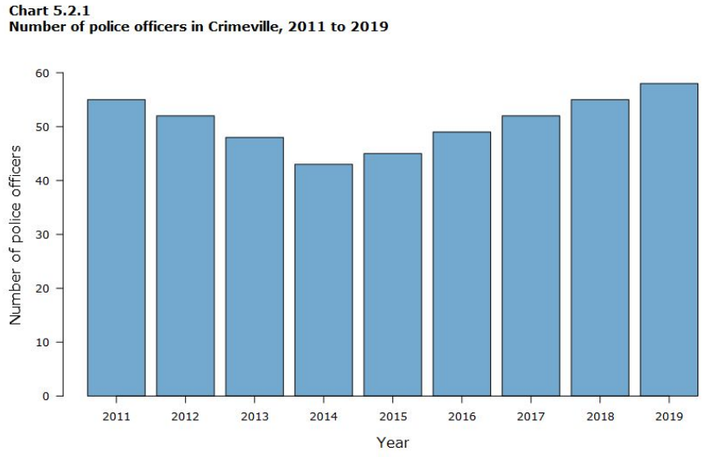


Count



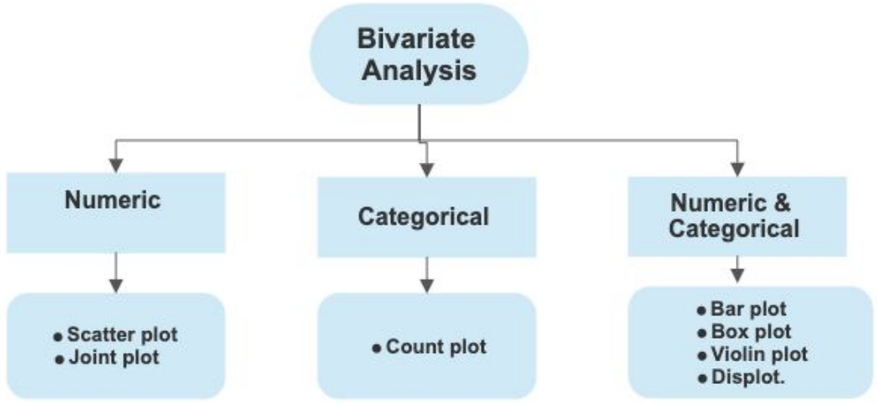
Count of variables

Bar

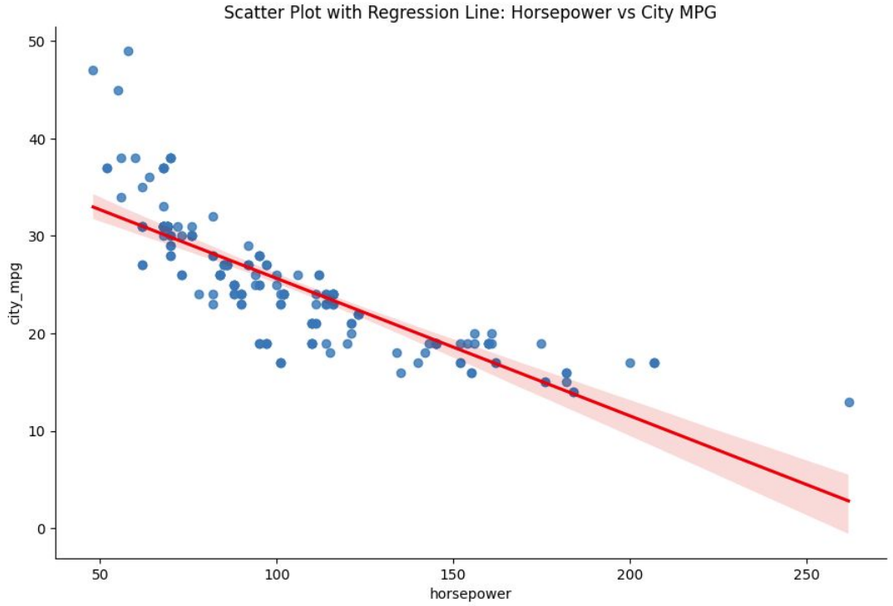


Count of variables

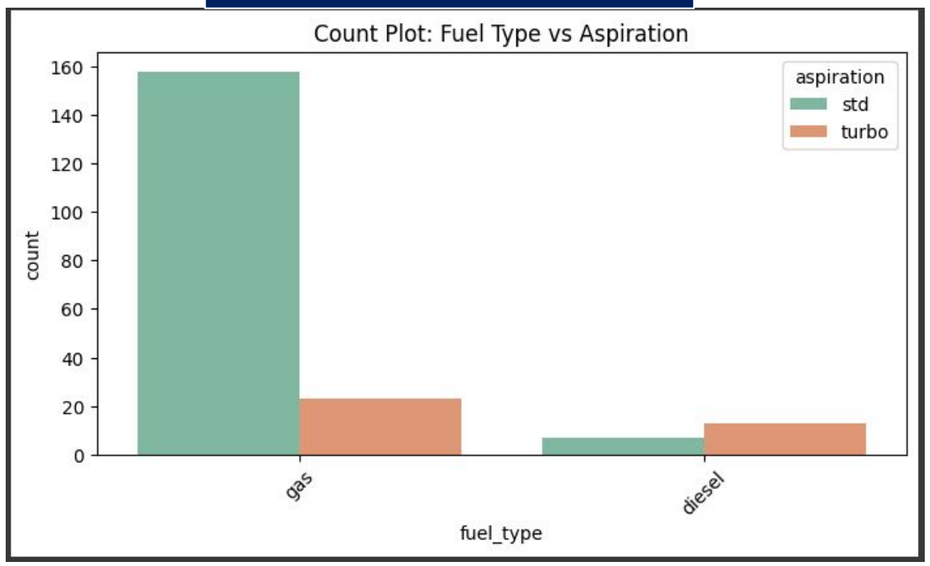
Bi-variate Analysis



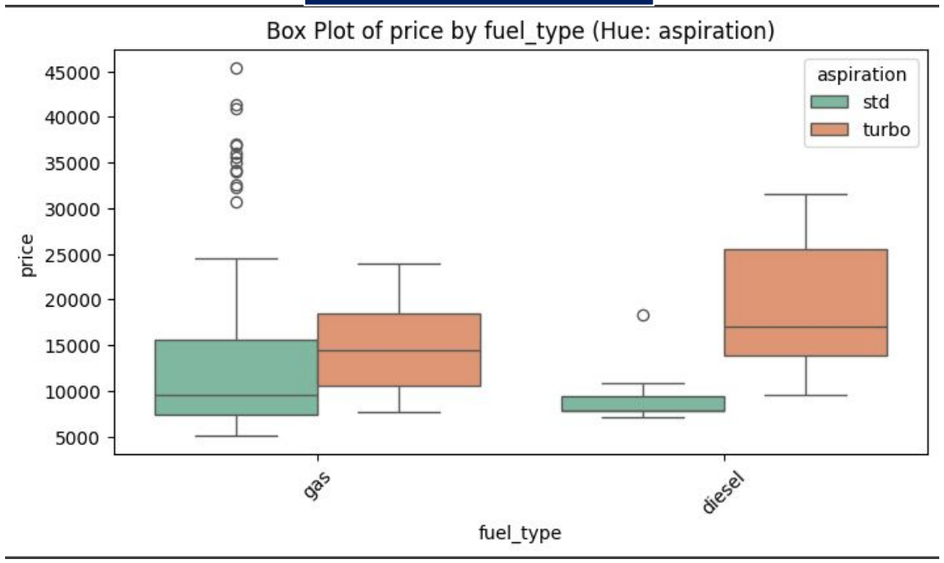
Scatter Plot (N-N)



Count Plot (C-C + Hue)



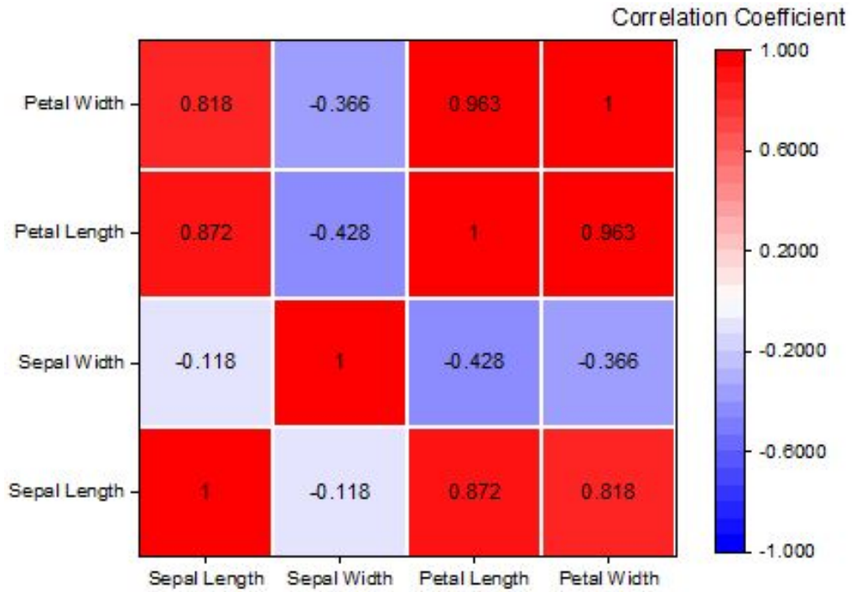
Box Plot (N-C + Hue)



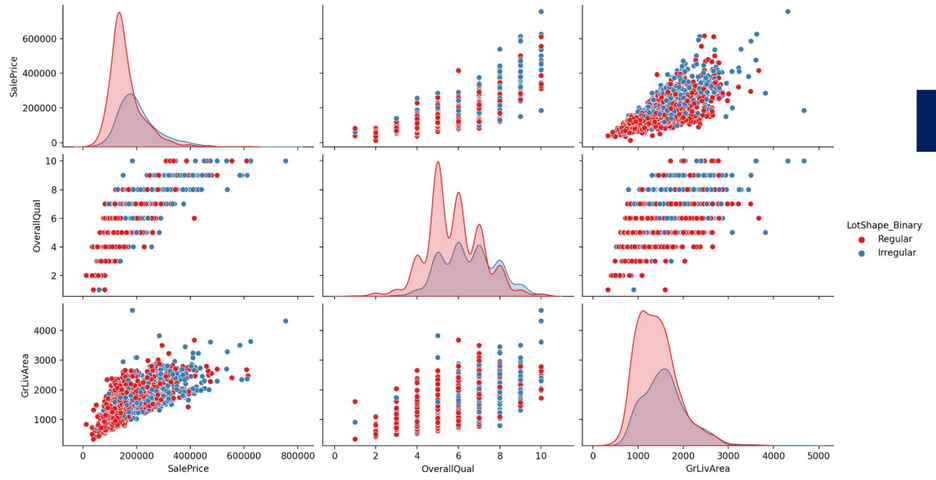
Multivariate Analysis

Multivariate Analysis

- Barplot
- Pairplot.
- Heatmap.
- Histogram

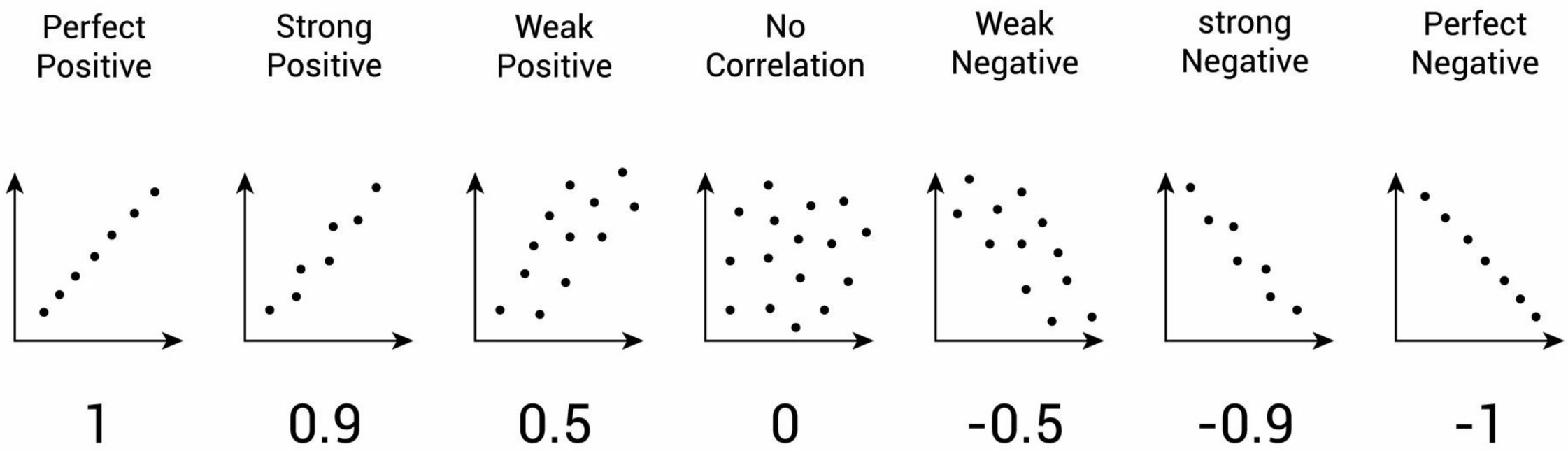


Heat Map



Pair Plot

Plot types - Correlation (Multivariate Numerical)- Heatmap



https://www.simplypsychology.org/correlation.html

Python for Data Science - Exploratory Data Analysis (EDA)_Part_3 - *Analysis of Graphs (Very Critical) - Your Assignment and Quiz is going to have a lot of question on the analysis*

Dataset - insurance.csv - Students will create notebook in class