

MSDS610 Fall 2022 - Exploratory Data Analysis

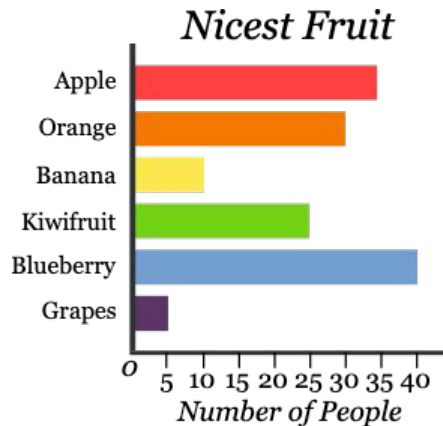
Ensun Pak
Abhradeep Mukherjee
Stephen Louie





What is EDA?

Technique used in Data Science used to analyze and investigate data sets and summarize their main characteristics, often employing data visualization methods.





Why EDA?

Due to the high volume of data that can be contained in a single file, EDA acts as an intermediary between asking a question about a business or operation and the development of a hypothesis.

- What is the distribution of of the data? Does the median align with the average? Does the data have any outliers?





How is EDA done?

Python

- an interpreted, object-oriented programming language with high-level, built-in data structures.
- The modules and prebuilt functions are ideal for creating visualizations such as histograms, box plots, stem and leaf plots.