

**REPORT
OF
DATA VISUALIZATION ASSIGNMENT USING
PYTHON**

PROJECT OF DATA VISUALIZATION

BY:

Name: Madipadige Saicharan

Email: madipadigesaicharan@gmail.com

SUBMITTED TO ARTIFICIAL NEURONS.AI

ABSTRACT

Data visualization is a brand-new and exciting area in computer science. In order to extract patterns, trends, and relationships from datasets, computer graphic effects are used. In this essay, we first familiarize ourselves with data visualization and topics that are related to it, after which we will examine some common data visualization algorithms. We shall talk about multidimensional data visualization to get deeper into it. We introduce a new methodology to perform four dimensional data visualization using a mix of some established techniques. We also discuss some related problems and explanations, as well as an optional programmed project plan.

INTRODUCTION

This dataset needs to be analyzed and visualized using python modules like: pandas, pandas-profiling, numpy, seaborn, matplotlib, etc. The given dataset consists of 4 categories: Date, Description, Unit (a unit of 1701 calories that is used in nutritional research), and Place, indicating that it is a Food Description dataset.

This project's goal is to use the provided dataset to produce a general summary of the food description and availability at a specific date and place.

INSIGHTS FROM DATA VISUALIZATION

Overview

OverviewAlerts4Reproduction

Dataset statistics

Number of variables	4
Number of observations	103
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	4
Duplicate rows (%)	3.9%
Total size in memory	4.0 KiB
Average record size in memory	40.0 B

Variable types

Categorical	3
Numeric	1

Variables

DATE

Categorical

HIGH CORRELATION

Distinct	33
Distinct (%)	32.0%
Missing	0
Missing (%)	0.0%
Memory size	1.6 KiB

Thursday, May 19, 202210

Thursday, April 21, 20229

Tuesday, June 21, 20227

Wednesday, April 20, 20226

Tuesday, May 17, 20225

Other values (28)66

Toggle details

OverviewCategoriesWordsCharacters

Length

Max length	25
Median length	24
Mean length	22.01941748
Min length	20

Characters and Unicode

Total characters	2268
Distinct characters	32
Distinct categories	5?
Distinct scripts	2?
Distinct blocks	1?

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	12?
Unique (%)	11.7%

Sample

1st row	Friday, April 1, 2022
2nd row	Friday, April 1, 2022
3rd row	Sunday, April 3, 2022
4th row	Tuesday, April 5, 2022
5th row	Wednesday, April 6, 2022

DESCRIPTION

Categorical

Distinct	9
Distinct (%)	8.7%
Missing	0
Missing (%)	0.0%
Memory size	1.6 KiB



Toggle details

Overview

Categories

Words

Characters

Length

Max length	9
Median length	6
Mean length	5.368932039
Min length	4

Characters and Unicode

Total characters	553
Distinct characters	24
Distinct categories	3 ?
Distinct scripts	2 ?
Distinct blocks	1 ?

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	1 ?
Unique (%)	1.0%

Sample

1st row	Beef
2nd row	Fish
3rd row	Beef
4th row	Prawns
5th row	Dates

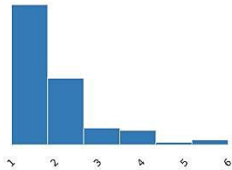
UNIT

Real number (ℝ_{≥0})

HIGH CORRELATION

Distinct	6
Distinct (%)	5.8%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	1.718446602

Minimum	1
Maximum	6
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	1.6 KiB



Toggle details

Statistics

Histogram

Common values

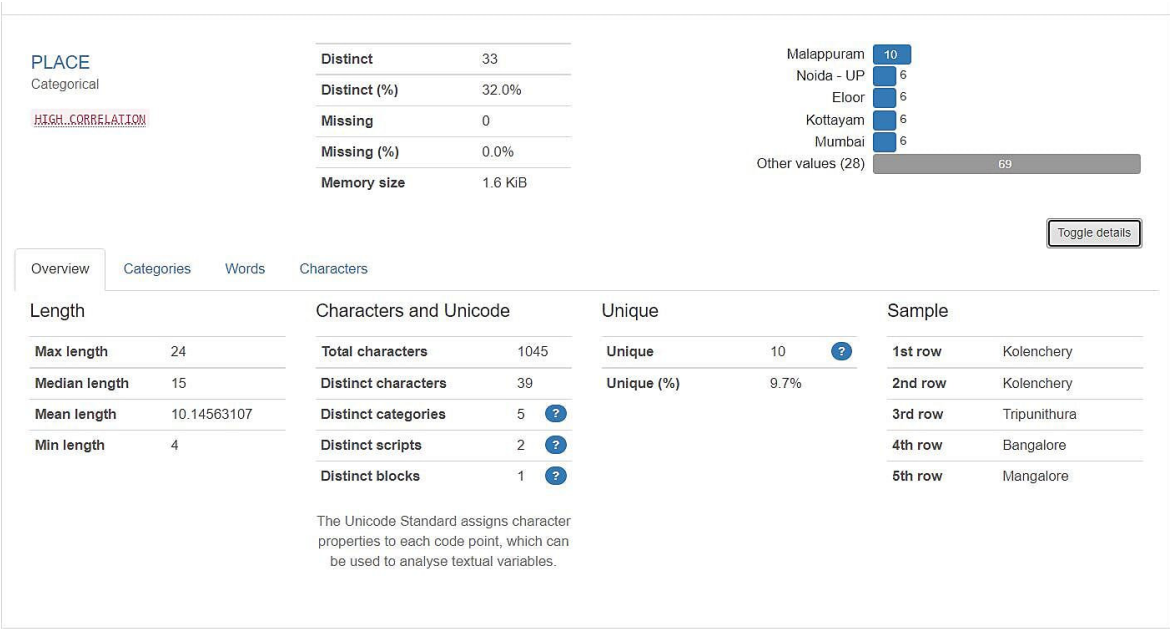
Extreme values

Quantile statistics

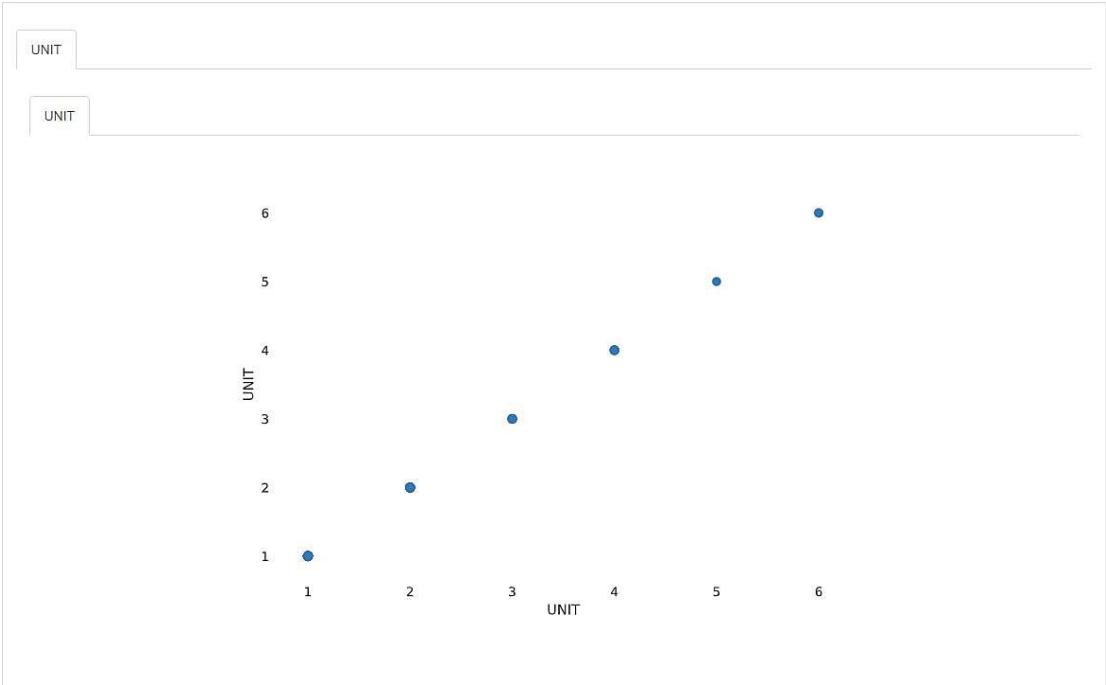
Minimum	1
5-th percentile	1
Q1	1
median	1
Q3	2
95-th percentile	4
Maximum	6
Range	5
Interquartile range (IQR)	1

Descriptive statistics

Standard deviation	1.097389735
Coefficient of variation (CV)	0.6385940264
Kurtosis	3.974973622
Mean	1.718446602
Median Absolute Deviation (MAD)	0
Skewness	1.944711372
Sum	177
Variance	1.20426423
Monotonicity	Not monotonic



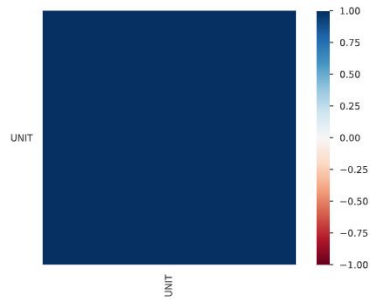
Interactions



Correlations

Spearman's ρ Pearson's r Kendall's τ Cramér's V (ϕ_c) Phik (ϕ_k)

Toggle correlation descriptions

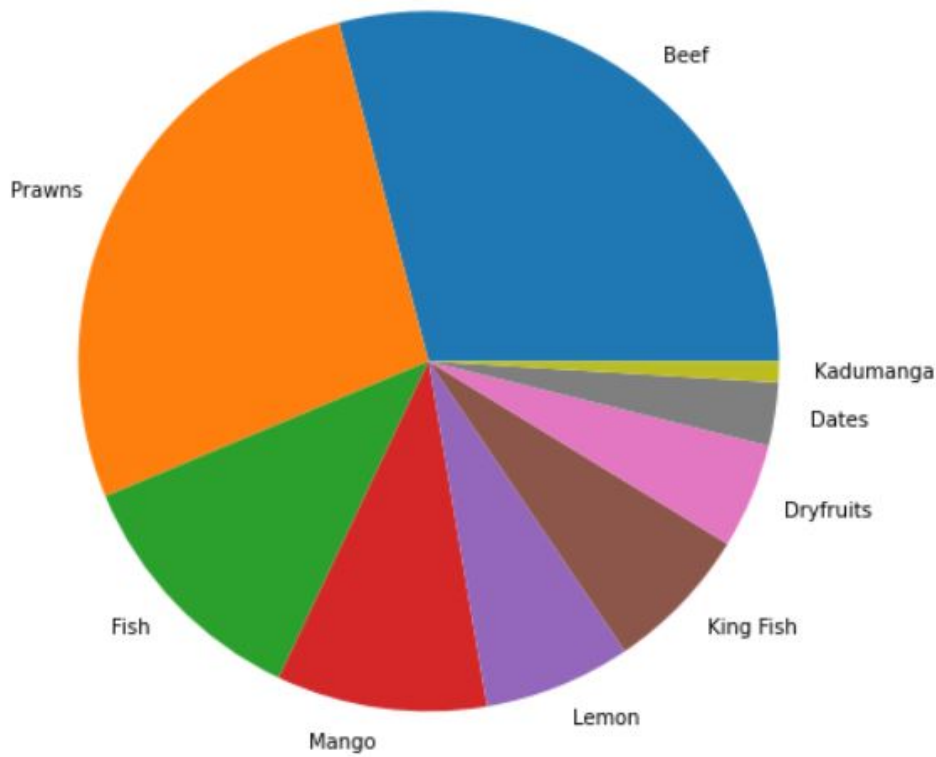


Spearman's ρ

The Spearman's rank correlation coefficient (ρ) is a measure of monotonic correlation between two variables, and is therefore better in catching nonlinear monotonic correlations than Pearson's r . It's value lies between -1 and +1, -1 indicating total negative monotonic correlation, 0 indicating no monotonic correlation and 1 indicating total positive monotonic correlation.

To calculate ρ for two variables X and Y , one divides the covariance of the rank variables of X and Y by the product of their standard deviations.

Food_Description



What has done?

Given dataset consists of null and duplicate values either we need to replace or remove the null values in order to maintain the data quality given dataset contains one numerical variable and three are categorical variables. So, I used python modules for cleaning the data and for exploratory data analysis and data visualization techniques like pie chart bar graph, heat map and for overview I used pandas-profiling which gives the complete information of dataset in html format.

What else can be done?

Given that the dataset only has four columns, we can add a vegetable food item in the Description Column. A fifth column can be added namely food pricing, which can include a description of the food and their price, and a sixth column called nutrients which includes carbohydrates, fats, proteins, fibers, vitamins, and water can be added and other column named storage temperature to make sure the food is safe to avoid deterioration.

Why was it done?

It assists in organizing material into a more comprehensible structure, eliminating ambiguous information, and emphasizing relevant details. It can organize the complicated data in easier and understandable format to describe the data in a graphical representation information and given data format.

Conclusion

Finally, I would like to say that employing data visualization can be effective and capable of portraying the larger datasets in a graphical manner that can be seen visually. I understand how to use python libraries like matplotlib and seaborn API in this assignment to make stunning visualizations for examining the relationship between the variables.