# \*\*Review of tips dataset\*\*

## \*\*Table of contents\*\*

\* [Introduction](#Introduction)

\* [Setup](#Setup)

\* [What is the tips dataset?](#What-is-the-tips-dataset?)

\* [Some statistics on the dataset](#Some-statistics-on-the-dataset)

\* [Description 1](#Description-1)

\* [Skewness and kurtosis 1](#Skewness-and-kutrosis-1)

\* [Plotting statistics](#Plotting-statistics)

\* [Covariance](#Covariance)

\* [Correlation](#Correlation)

\* [Description discussion](#Description-discussion)

\* [Regression analysis: Table bill vs tips](#Regression-analysis:-Table-bill-vs-tips)

\* [Best fit and mean fit lines](#Best-fit-and-mean-fit-lines)

\* [Regression discussion](#Regression-discussion)

\* [Some other analysis](#Some-other-analysis)

\* [Time of dinner](#time-of-dinner)

\* [Smoker](#Smoker)

\* [Description 2](#Description-2)

\* [Covariance and correlation](#Covariance-and-correlation)

\* [Skewness and kurtosis 2](#Skewness-and-kurtosis-2)

\* [Plotting other analysis](#Plotting-other-analysis)

\* [Discussion of other analysis](#Discussion-of-other-analysis)

\* [Further potential analysis](#Further-potential-analysis)

\*[Bibliography](#Bibliography)

## \*\*Introduction\*\*

In this notebook, it is intended to review the tips dataset, which can be found online. To do this, a number of approaches will be taken. The first step will be to review the data using standard descriptive statistics, such as mean, mode, median, quartile ranges, standard deviation, etc. Other information will also be gleaned, such as the skewness, and kurtosis of the data (this will be explained in the relevant section), as well as correlation and covariance between certain data sets.

A second step will be to conduct some analysis between the fields of total-bill and tip, to determine a relationship. This analysis will also include both a best fit, and a mean fit lines. These lines can be used for various reasons, and will be discussed in the regression analysis section.

The third section, two subsets of the data will be extricated from the data, and compared to each other. The two subsets of the data chosen for this analysis are the time that the diner at the location, as well as whether the diner was a smoker or not. Similar analysis will conducted with regards to these two data sets, as was conducted on the overall dataset. A discussion of the results can be found in the discussion of further analysis section.

A final section covers other potential analysis that could be undertaken, using the same dataset.

The tips dataset that was used in the below was sourced from [here](http://<https://raw.githubusercontent.com/mwaskom/seaborn-data/master/tips.csv>). Note the data here is the raw form, to make it easier to import into our code.

\*Note: in this text, a number of ideas, or explanations may be mentioned. These have all been sourced either from the works listed in the bibliography, or from the author’s own ideas. For the sake of legibility, unless a direct quote, comparison, or specific suggestion is used, it will not be cited in the standard academic (Harvard) manner.

### Setup

Before running any analysis of the data, it is necessary to import the data, and the libraries that we will be using. In this case, it is intended to import the following libraries:

- \*\*Numpy\*\*: To run some functions for analysis.

- \*\*Pandas\*\*: The dataset will be held in various pandas dataframes. The pandas library will also be used to conduct various analytical exercises.

- \*\*Seaborn\*\*: Will be used for various plotting functionality.

- \*\*Matplotlib.pyplot\*\*: For use in some graphing functionality.

Additionally, throughout this notebook, the main dataframe that will hold the dataset will be called \*tips\_ds\*. The below code is used to import the libraries and dataset.

It should be noted that the above code does not have any outputs. The confirmation that the code was successfully imported can be found at the end of [What is the tips dataset?](#What-is-the-tips-dataset?) (below).

## \*\*What is the tips dataset?\*\*

The tips dataset is a set of data that is often used to teach analytical skills, especially around the segregation of data, and analysis of a group, or a whole set. Additionally, it could be used for extrapolation of data, where the data in question is not included in the actual dataset. The data forms part of a tutorial for the Python Seaborn package, and is used to demonstrate how to conduct plotting of data in Python.

The tips dataset was collected in a restaurant. (Jain, 2017) suggests that the data was collected over a 2.5 month period, in a branch of a chain restaurant in California, USA. The data was collected by a server (waiter/waitress), and recorded each table they served, and included a variety of information regarding the table.

This information that was recorded fell into 7 different columns. These were:

- \*\*total-bill\*\*: the total bill for each meal (numerical values to two decimal places).

- \*\*tip\*\*: the tip received for each meal (numerical values to two decimal places).

- \*\*sex\*\*: gender of the table (either male or female. Doesn’t note if this was the gender of the payer, but would be safe to assume so).

- \*\*smoker\*\*: whether the table being served was in a smoking or non-smoking section (this was California in the 1990s after all)

- \*\*day\*\*: day of the week that the meal was being served. Covers only Thursday to Sunday.

- \*\*time\*\*: the time of day that the service was being conducted. Broken down as either lunch or dinner service.

- \*\*size\*\*: the number of people in the party (whole numerical value).

In total the collector made 244 entries into the dataset.

All the above data can be confirmed by Python.

## \*\*Some statistics on the dataset\*\*

#### Description 1

#### Skewness and kurtosis 1

#### Plotting statistics

#### Covariance

#### Correlation

#### Description discussion

## \*\*Regression analysis: Table bill vs tips\*\*

#### Best and mean fit lines

#### Regression discussion

## \*\*Some other analysis\*\*

### Time of dinner

### Smoker

#### Description 2

#### Covariance and correlation

#### Skewness and kurtosis 2

#### Plotting other analysis

#### Discussion of other analysis

## \*\*Further potential analysis\*\*

## \*\*Bibliography\*\*