

This member-only story is on us. [Upgrade](#) to access all of Medium.

◆ Member-only story

14 Data Visualization Plots of Seaborn

Tools for the cops who investigate data and extract information and trends from it.



Aayush Ostwal · Follow



Published in Towards Data Science · 7 min read · Sep 12, 2020



338



1



...

Open in app ↗



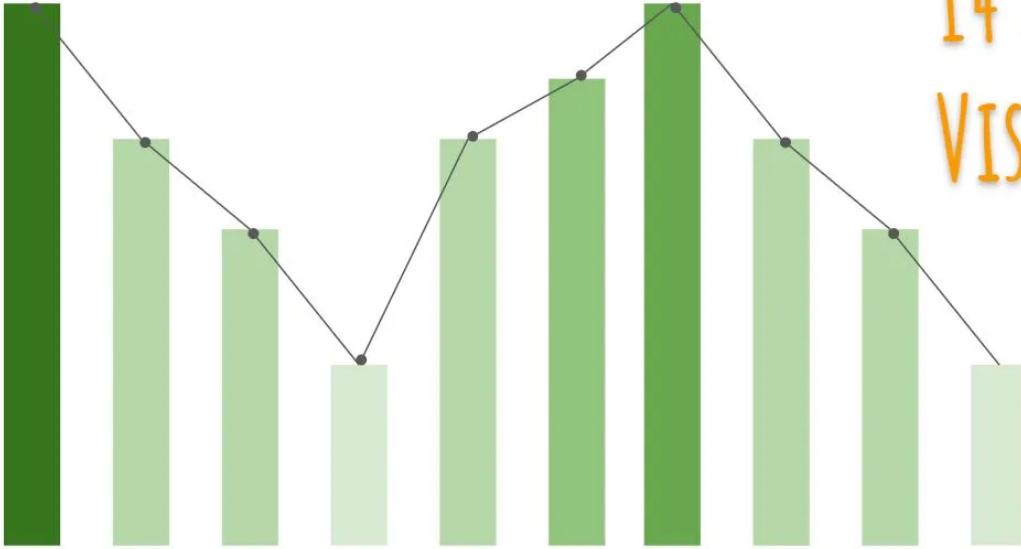
Search



Write



14 DATA VISUALIZATIONS



Green Colour represents **new beginnings** and **growth**. It also signifies **renewal** and **abundance**.

Data Visualization plays a very important role in Data mining. Various data scientist spent their time exploring data through visualization. To accelerate this process we need to have a well-documentation of all the plots.

Even plenty of resources can't be transformed into valuable goods without planning and architecture. Therefore I hope this article would provide you a good architecture of all plots and their documentation.

Content

1. *Introduction*
2. *Know your Data*
3. *Distribution Plots*
 - a. *Dist-Plot*
 - b. *Joint Plot*

c. Pair Plot

d. Rug Plot

4. *Categorical Plots*

a. Bar Plot

b. Count Plot

c. Box Plot

d. Violin Plot

5. *Advanced Plots*

a. Strip Plot

b. Swarm Plot

6. *Matrix Plots*

a. Heat Map

b. Cluster Map

7. *Grids*

a. Facet Grid

8. *Regression Plots*

Introduction

Seaborn is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics.

For the installation of Seaborn, you may run any of the following in your command line.

```
pip install seaborn  
conda install seaborn
```

To import seaborn you can run the following command.

```
import seaborn as sns
```

Know Your Data

The data set used in these plots is famous titanic data set (Fig. 1). Hereafter the data set is represented by the variable ‘df’.

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	Nan	S
1	2	1	Cumings, Mrs. John Bradley (Florence Briggs Th... Heikkinen, Miss. Laina	female	38.0	1	0	PC 17599 STON/O2. 3101282	71.2833 7.9250	C85 NaN	C S
2	3	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	26.0	0	0	113803	53.1000	C123	S
3	4	1	Allen, Mr. William Henry	male	35.0	1	0	373450	8.0500	Nan	S
4	5	0				0	0				

Fig. 1: Titanic Data set

Distribution Plots

These plots help us to visualize the **distribution** of data. We can use these plots to understand the mean, median, range, variance, deviation, etc of the data.

a. Distplot

- Dist plot gives us the histogram of the selected continuous variable.
- It is an example of a univariate analysis.

- We can change the number of bins i.e. number of vertical bars in a histogram

```
import seaborn as sns
sns.distplot(x = df['age'], bins = 10)
```

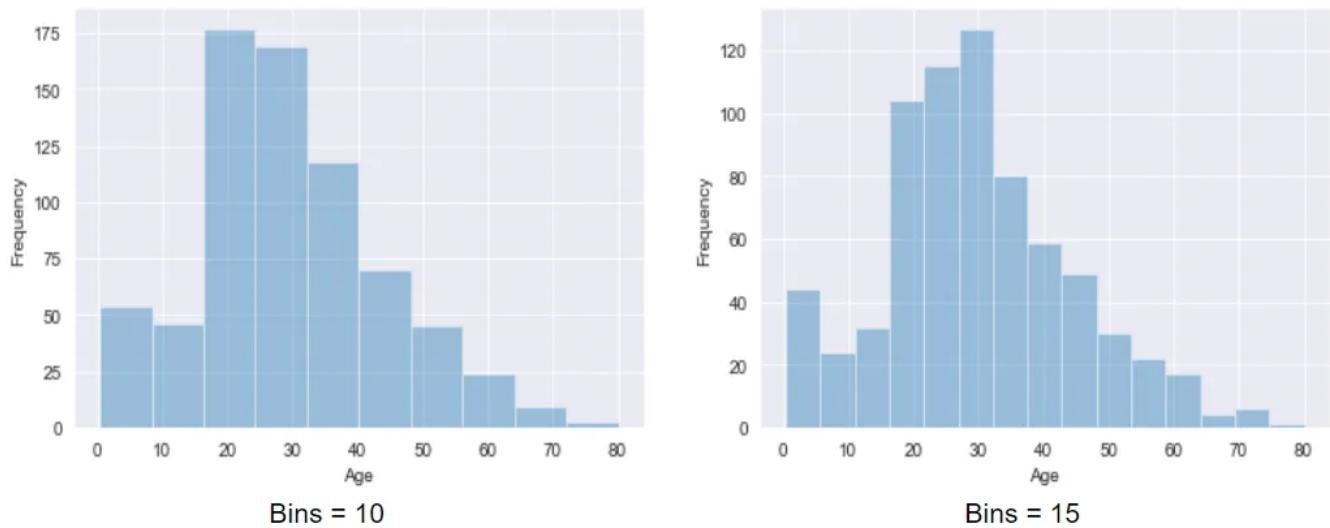


Fig. 2: Distribution Plot for 'Age' of Passengers.

- Here x-axis is the age and the y-axis displays frequency. For example, for bins = 10, there are around 50 people having age 0 to 10

b. Joint Plot

- It is the combination of the distplot of two variables.
- It is an example of bivariate analysis.
- We additionally obtain a scatter plot between the variable to reflecting their linear relationship. We can customize the scatter plot into a hexagonal plot, where, more the color intensity, the more will be the number of observations.

```

import seaborn as sns
# For Plot 1
sns.jointplot(x = df['age'], y = df['Fare'], kind = 'scatter')

# For Plot 2
sns.jointplot(x = df['age'], y = df['Fare'], kind = 'hex')

```

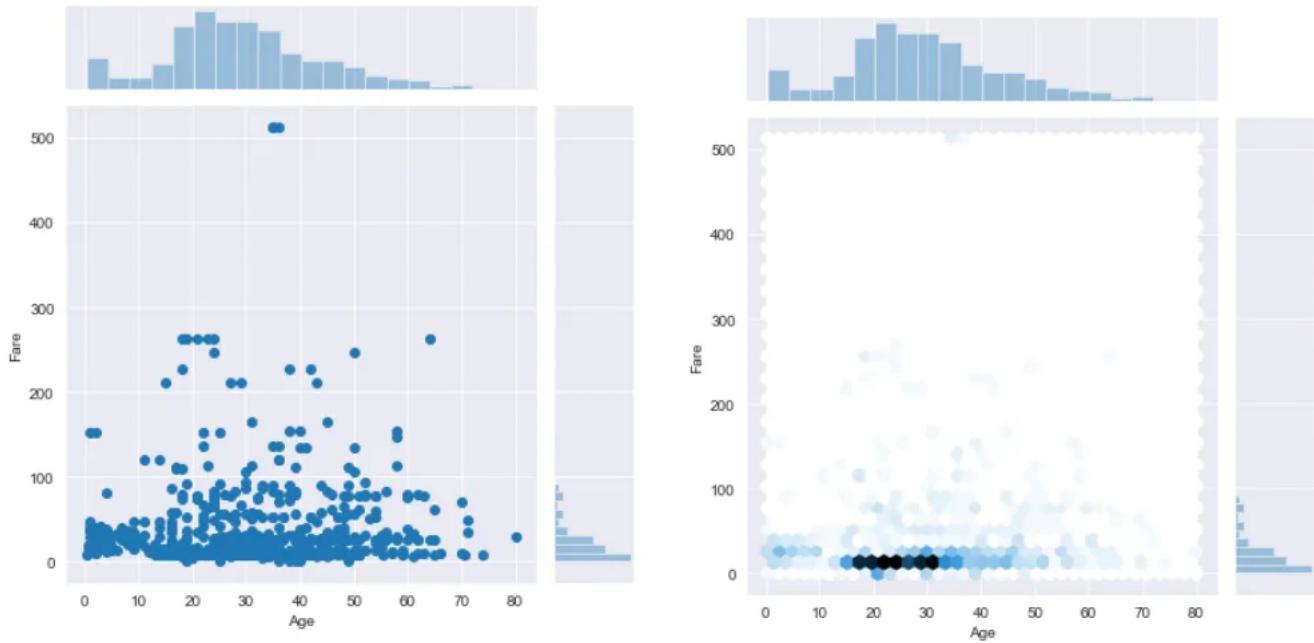


Fig. 3: Joint plots between 'Age' and 'Fare'

- We can see that there no appropriate linear relation between age and fare.
- kind = 'hex' provides the hexagonal plot and kind = 'reg' provides a regression line on the graph.

c. Pair Plot

- It takes all the numerical attributes of the data and plot pairwise scatter plot for two different variables and histograms from the same variables.

```
import seaborn as sns
```

`sns.pairplot(df)`



Fig. 4: Pair Plot of the titanic Data set

d. Rug Plot

- It draws a dash mark instead of a uniform distribution as in distplot.

- It is an example of a univariate analysis.

```
import seaborn as sns  
sns.rugplot(x = df['Age'])
```

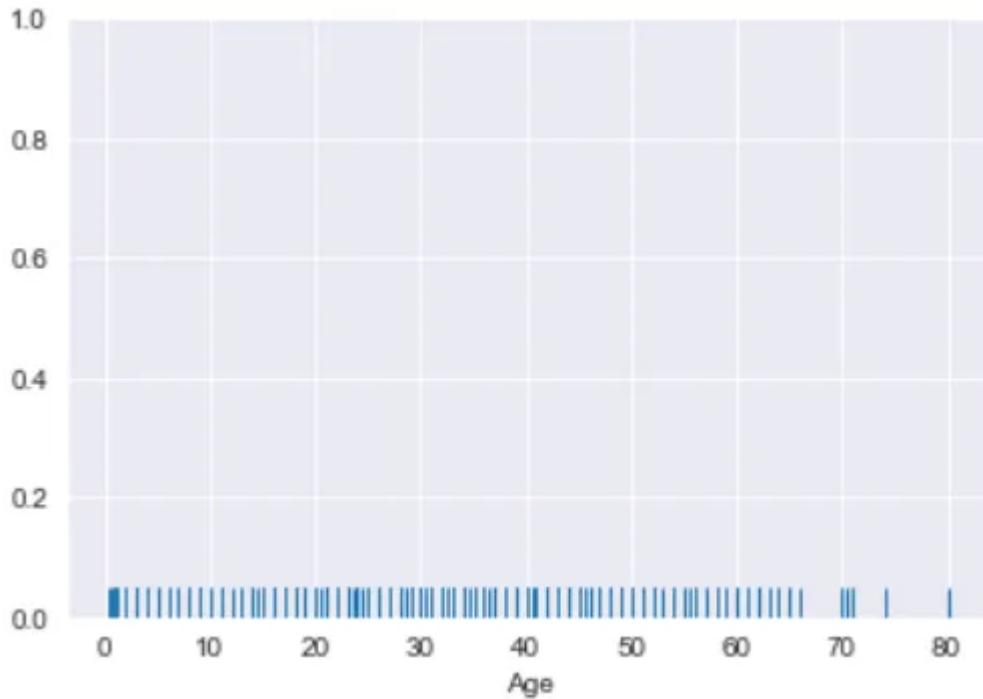


Fig. 5: Rug Plot for 'Age' of Passengers

Categorical Plots

These plots help us understand the **categorical** variables. We can use them for both univariate and bivariate analysis.

a. Bar Plot

- It is an example of bivariate analysis.

- On the x-axis, we have a categorical variable and on the y-axis, we have a continuous variable.

```
import seaborn as sns
sns.barplot(x = df['Sex'], y = df['Fare'])
```

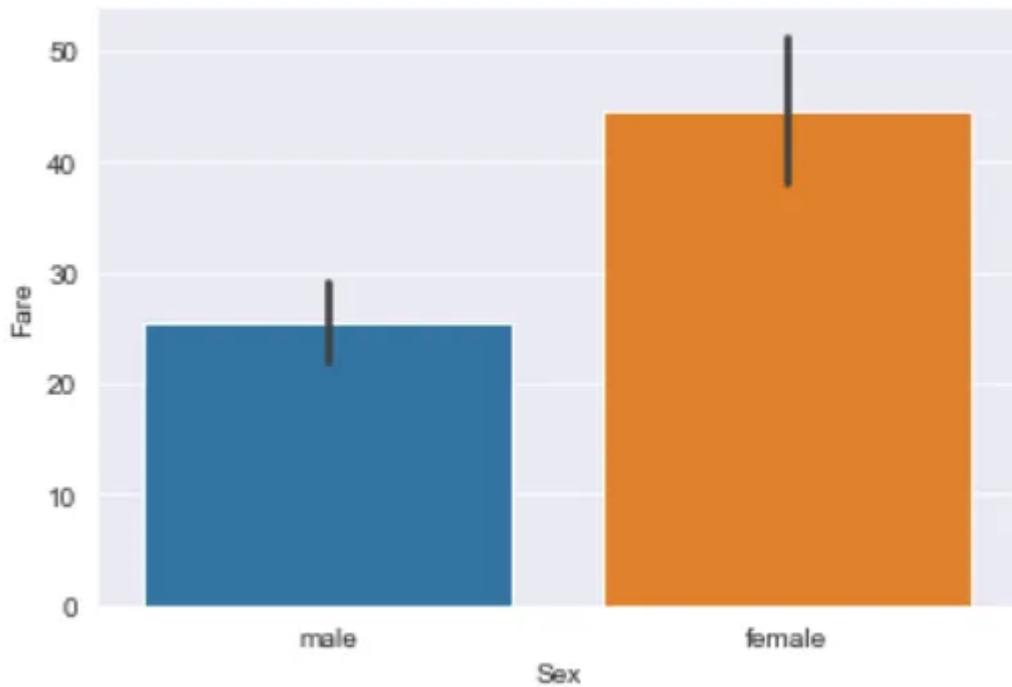


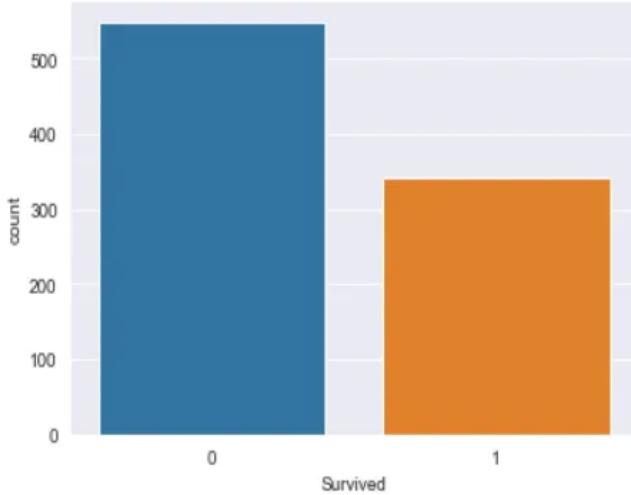
Fig. 6: Bar plot for 'Fare' and 'Sex'

- We can infer that the average fare is higher for females than males.

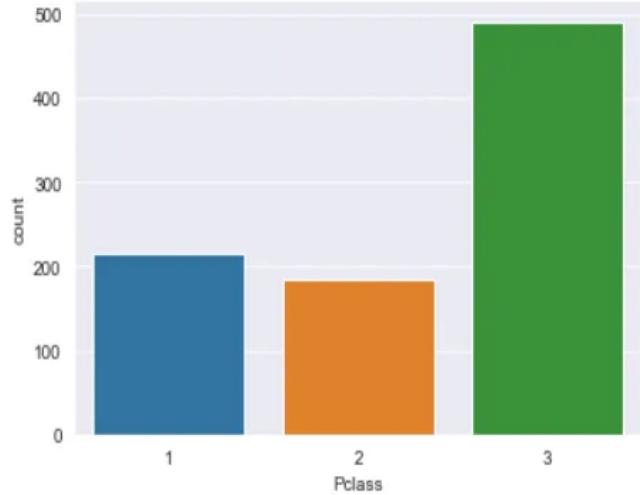
b. Count Plot

- It counts the number of occurrences of categorical variables.
- It is an example of a univariate analysis.

```
import seaborn as sns  
sns.countplot(df['Pclass'])
```



342 persons survived and 549 died



491 person belongs to class 3, 216 to class 1 and 184 to class 2

Fig. 7: Count Plot for Survived and 'P-class'.

c. Box Plot

- It is a **5 point summary plot**. It gives the information about the maximum, minimum, mean, first quartile, and third quartile of a continuous variable. Also, it equips us with knowledge of outliers.
- We can plot this for a single continuous variable or can analyze different categorical variables based on a continuous variable.

```
import seaborn as sns  
#For plot 1  
sns.countplot(df['Pclass'])  
  
#For plot 2  
sns.boxplot(y = df['Age'], x = df['Sex'])
```

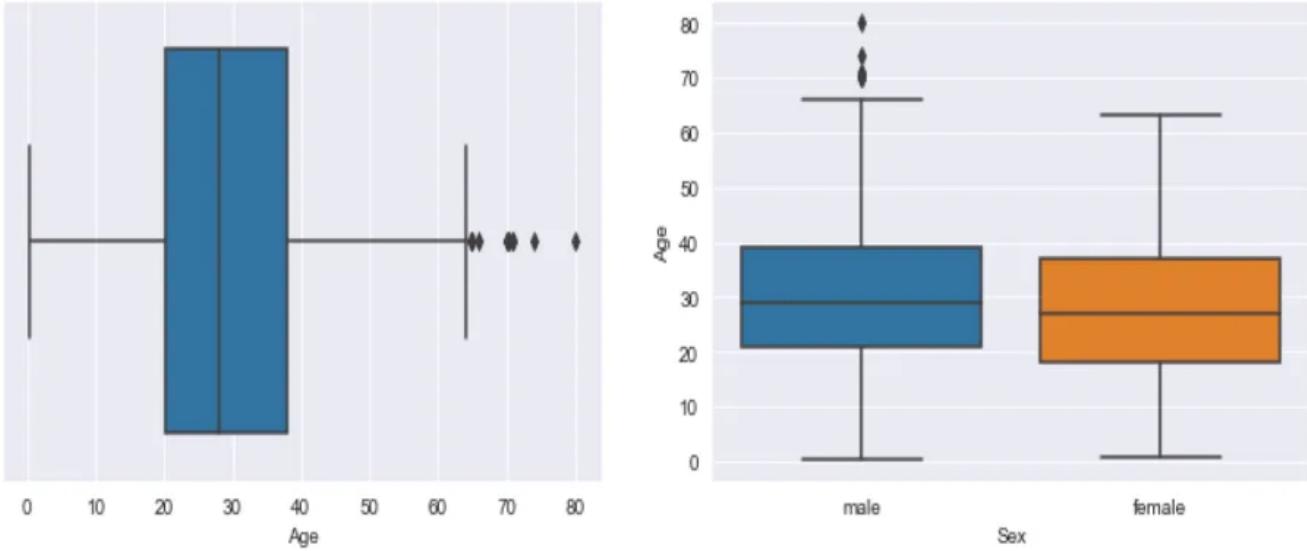


Fig.8: a) Box plot of 'Age', b) Box plot of different categories in 'sex' for 'Age'

d. Violin Plot

- It is similar to the Box plot, but it gives supplementary information about the distribution too.

```
import seaborn as sns
sns.violinplot(y = df['Age'], x = df['Sex'])
```

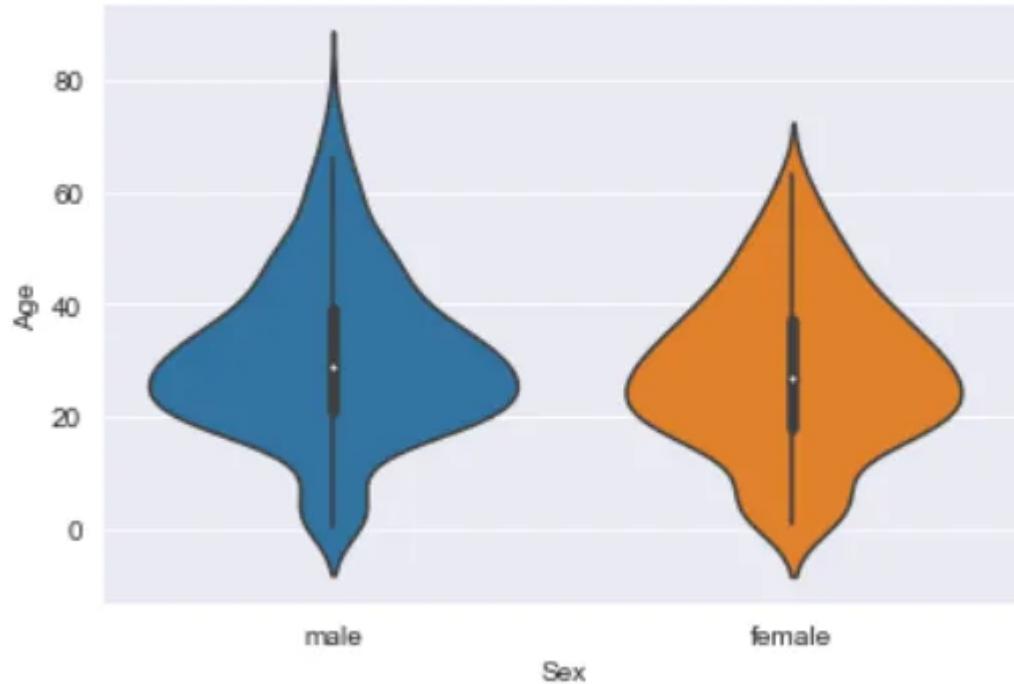


Fig. 9: Violin Plot between 'Age' and 'Sex'

Advanced Plots

As the name suggests, they are advanced because they ought to fuse the distribution and categorical encodings.

a. Strip Plot

- It's a plot between a continuous variable and a categorical variable.
- It plots as a scatter plot but supplementarily uses categorical encodings of the categorical variable.

```
import seaborn as sns
sns.stripplot(y = df['Age'], x = df['Pclass'])
```

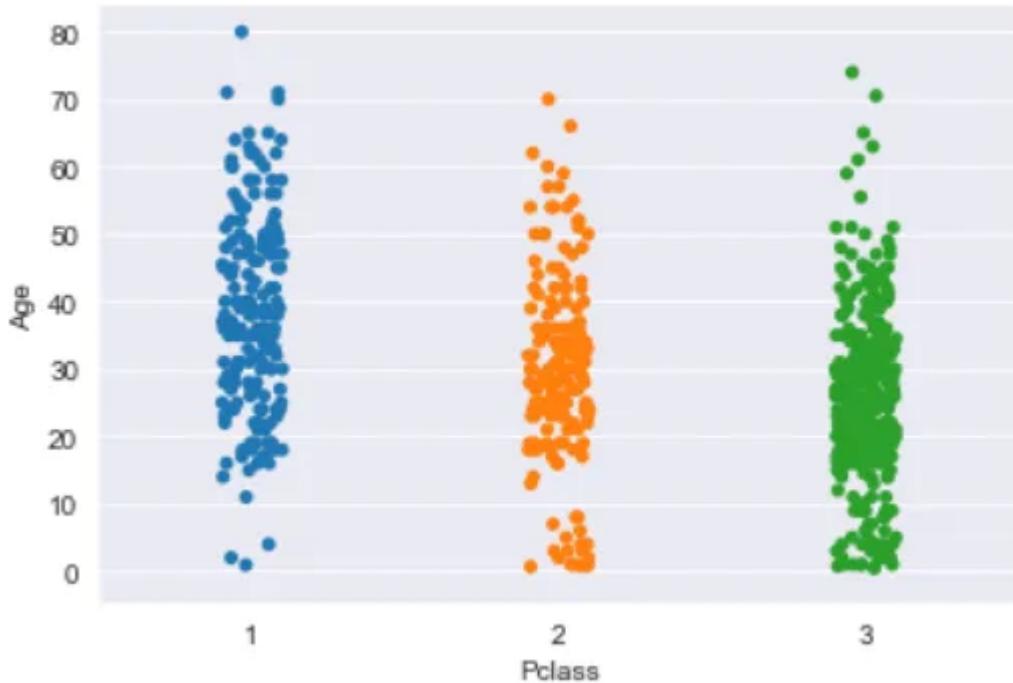


Fig.10: Strip Plot between 'Age' and 'P-class'

- We can observe that in class 1 and class 2, children around 10 years are not present and the people having age above 60 are mostly accommodated in class 1.
- Usually, these types of observations are used to impute missing values.

b. Swarm Plot

- It is the combination of a strip plot and a violin plot.
- Along with the number of data points, it also provides their respective distribution.

```
import seaborn as sns
sns.swarmplot(y = train['Age'], x = train['Pclass'])
```

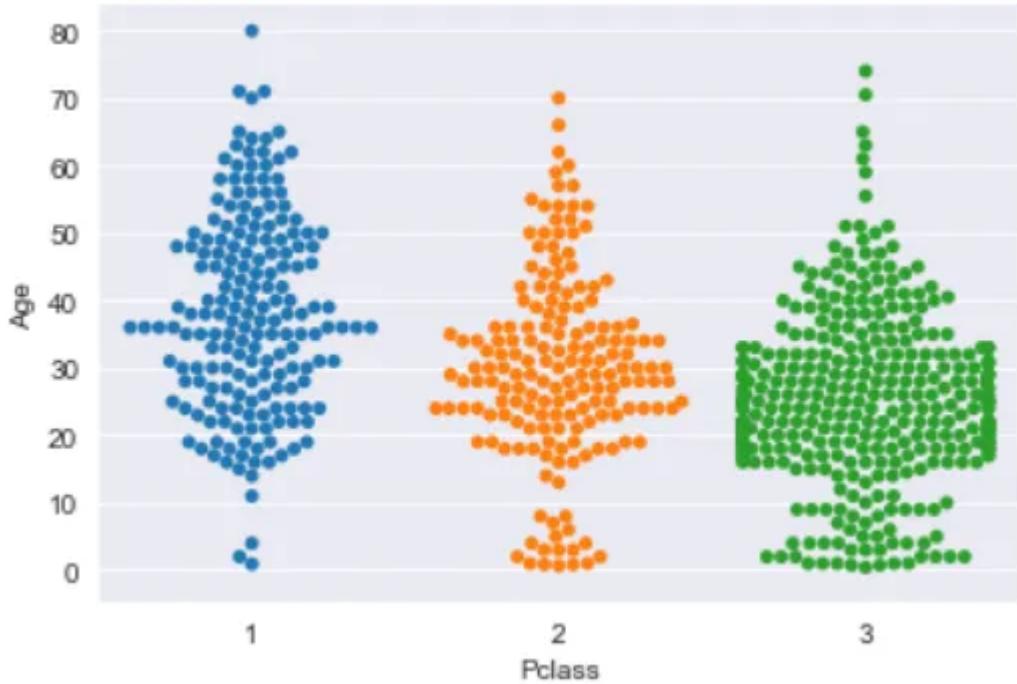


Fig. 11: Swarm Plot between 'Age' and 'P-class'

Matrix Plots

These are the special types of plots that use **two-dimensional matrix data** for visualization. It is difficult to analyze and generate patterns from matrix data because of its large dimensions. So, this makes the process easier by providing color coding to matrix data.

a. Heat Map

- In the given raw dataset 'df', we have seven numeric variables. So, let us generate a correlation matrix between these seven variables.

```
df.corr()
```

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
PassengerId	1.000000	-0.005007	-0.035144	0.036847	-0.057527	-0.001652	0.012658
Survived	-0.005007	1.000000	-0.338481	-0.077221	-0.035322	0.081629	0.257307
Pclass	-0.035144	-0.338481	1.000000	-0.369226	0.083081	0.018443	-0.549500
Age	0.036847	-0.077221	-0.369226	1.000000	-0.308247	-0.189119	0.096067
SibSp	-0.057527	-0.035322	0.083081	-0.308247	1.000000	0.414838	0.159651
Parch	-0.001652	0.081629	0.018443	-0.189119	0.414838	1.000000	0.216225
Fare	0.012658	0.257307	-0.549500	0.096067	0.159651	0.216225	1.000000

Fig. 12: Correlation matrix

- It seems very difficult to read every value even though there are only 49 values. The intricacy intensifies as we traverse towards thousands of features.

So, let us try to implement some color coding and see how easy the interpretation becomes.

```
sns.heatmap(df.corr(), annot = True, cmap = 'viridis')
```

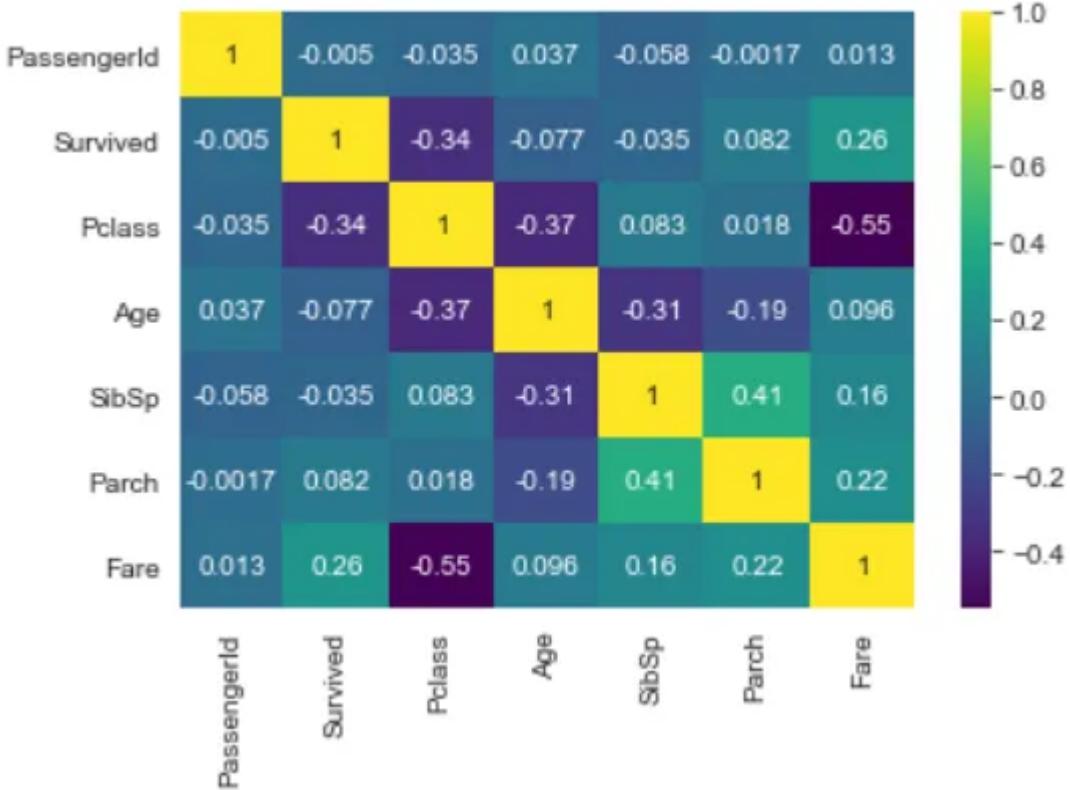


Fig. 13: Heat Map of the correlation matrix of the titanic data set.

- The same matrix is now articulating more information.
- Another very obvious example is to use heatmaps to understand the missing value patterns. In Fig. 14, the yellow dash represents a missing value, hence it makes our tasks more effortless to identify the missing values.

```
sns.heatmap(df.isnull(),yticklabels=False,cbar=False,cmap='viridis')
```

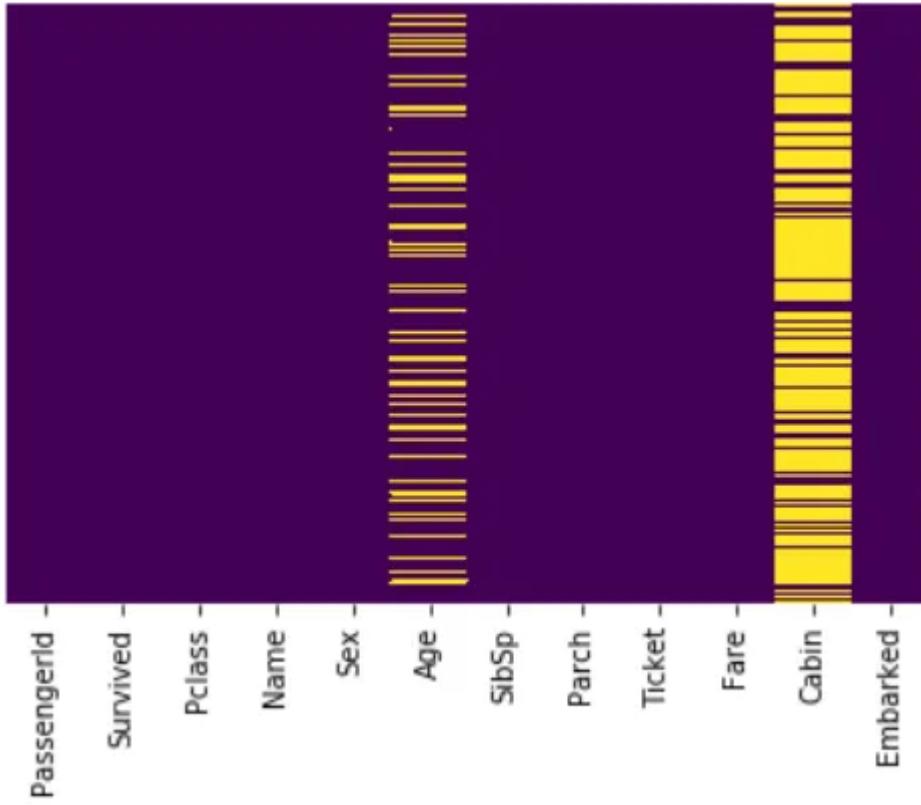


Fig. 14: Heat Map for missing values in titanic data.

b. Cluster Map

- If we have a matrix data and want to group some features according to their similarity, cluster maps can assist us. Once, have a look at the heat map (Fig. 13) and then look at the cluster map (Fig. 15).

```
sns.clustermap(tran.corr(), annot='True',cmap='viridis')
```

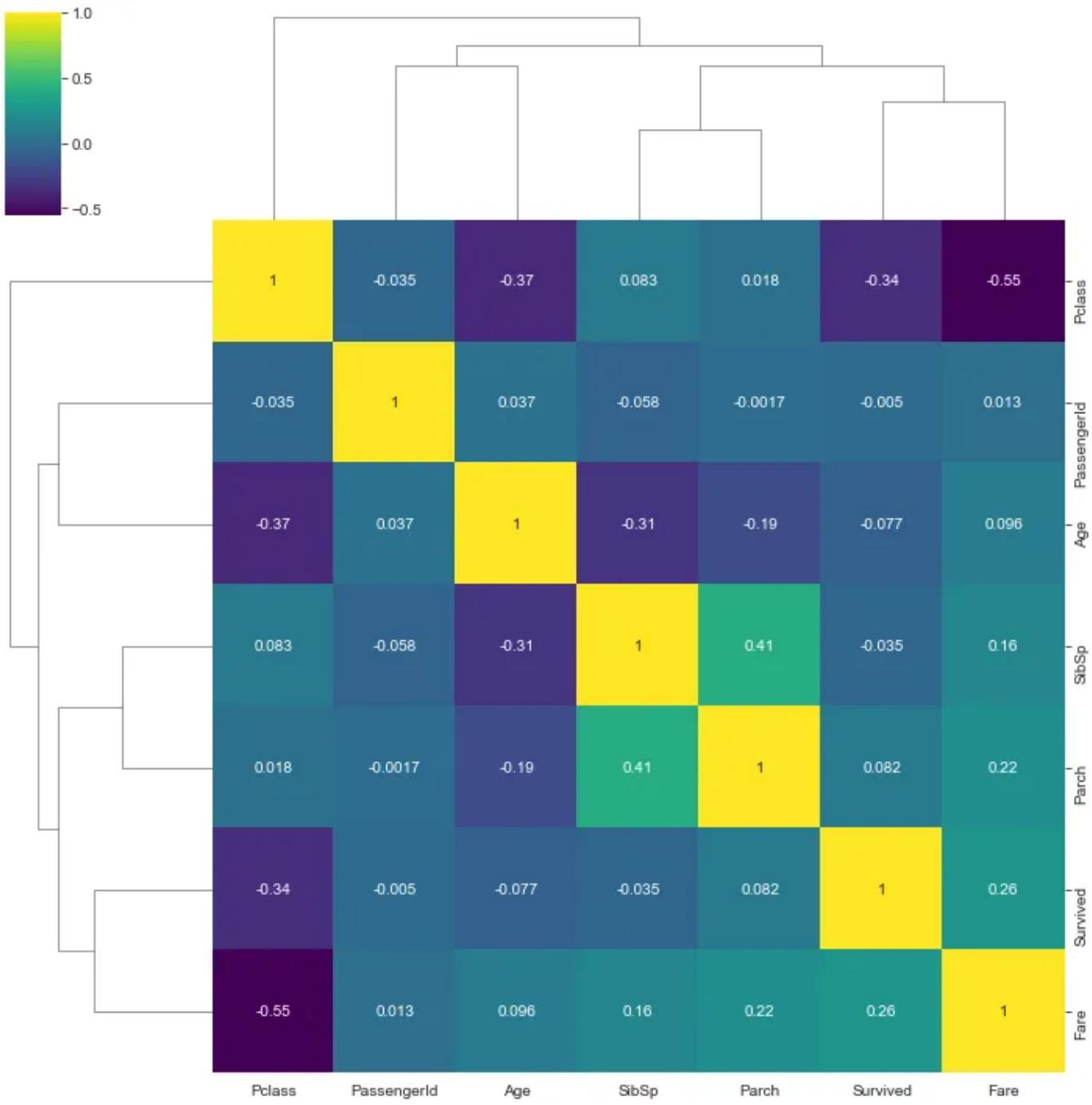


Fig. 15: Cluster map for correlation matrix of titanic data

- The x-label and y-label are the same but they harmonized differently. That is because they are grouped according to their similarity.
- The flow-chart like structure at the top and left describe their degree of similarity.
- Cluster maps use **Hierarchical clustering** to form different clusters.

Grids

Grid plots provide us more control over visualizations and plots various assorted graphs with a single line of code.

a. Facet Grid

- Suppose we want to plot the age distribution of males and females in all the three classes of tickets. Hence, we would be having in a total of 6 graphs.

```
sns.FacetGrid(train, col = 'Pclass', row = 'Sex').map(sns.distplot,  
'Age')
```

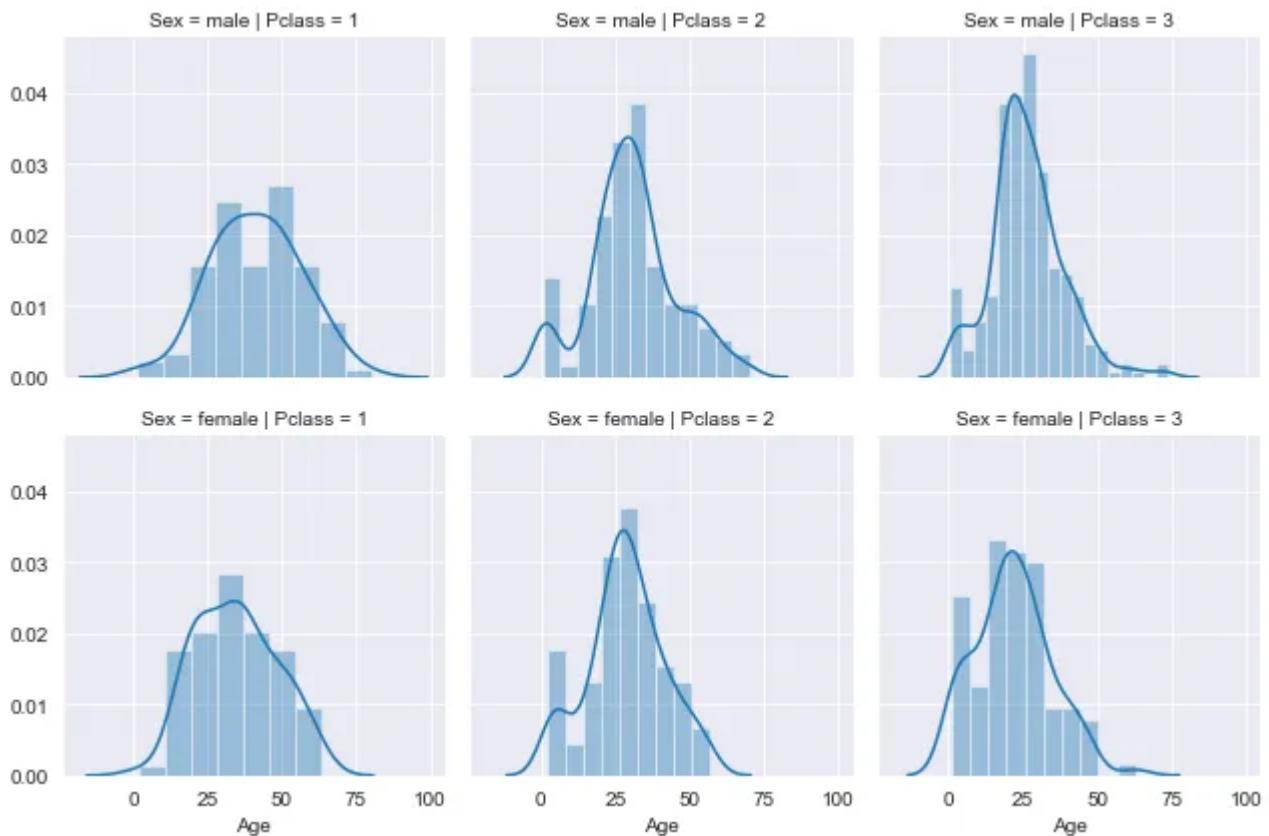


Fig. 16: Distribution plot of 'Age' for classes of 'Sex' and 'P-class'

- The Facet grids provide very clear graphs as per requirements.
- `sns.FacetGrid(col = 'col', row = 'row', data = data)` provides an empty grid of all unique categories in the *col* and *row*. Later, we can use different plots and common variables for peculiar variations.

Regression Plot

This is a more advanced statistical plot that provides a scatter plot along with a linear fitting on the data.

```
sns.lmplot(x = 'Age', y = 'PassengerId', data = df, hue = 'Sex')
```

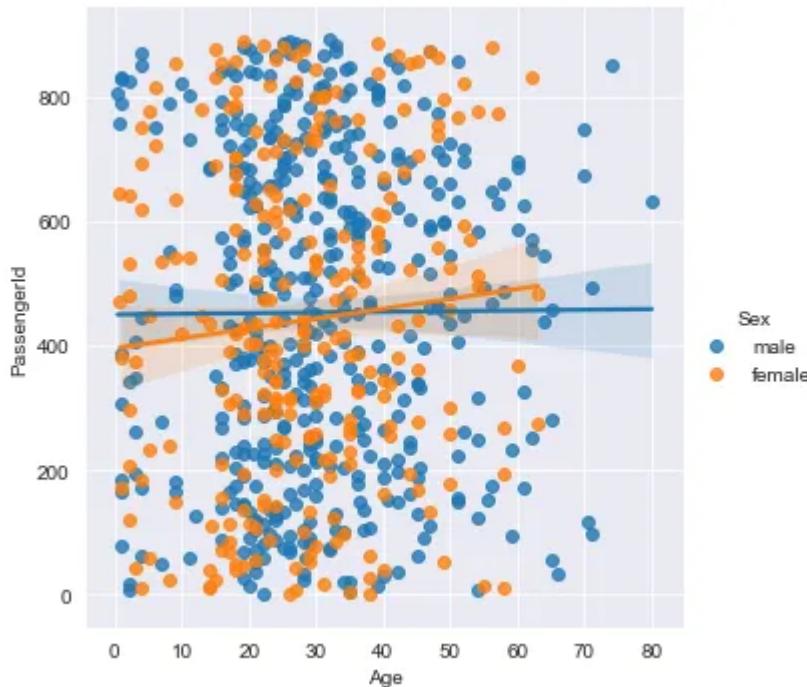


Fig. 17: Regression Plot between Age and Passenger ID for males and females. | **Disclaimer:** There is so the significance of regressing age and passenger id. It is just the purpose of understanding visualization.

Fig. 17 displays the linear regression fitting between Passenger ID and Age for both males and females.

Wrap Up

In this article, we have seen 14 different visualization techniques using seaborn.

I believe data visualization enhances our understanding and potential for interpreting data. It gives us more satisfying skills to represent data, impute missing values, identify outliers, detect anomalies, and a lot more.

Data Analysts are like cops that need to interrogate data and extract information via them. It is extremely necessary to have optimistic tools to do the job. Therefore, I hope this article would serve you as a tool for interrogating your data.

For the Guide for Exploratory data analysis, visit-

The EDA Theoretical Guide

A complete guide to exploratory data analysis

towardsdatascience.com

For such content related to data science, machine learning, programming, please visit my youtube channel.

Aayush Ostwal

Namaste!!! 🙏🙏 This Channel is for those who want to learn Data Science and Machine learning. I have spent a lot of...

Happy Learning!

Data Science

Machine Learning

Data Visualization

AI

Data Analysis



Written by Aayush Ostwal

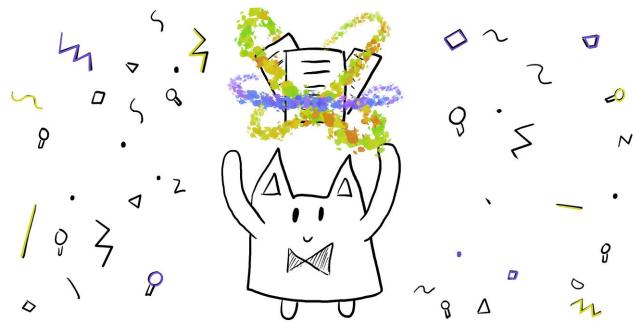
160 Followers · Writer for Towards Data Science

Follow



AI Engineer at Qure.ai | Enthusiastic ML practitioner | IIT Kanpur | Drama Lover | [Subscribe](#)
https://www.youtube.com/channel/UCqq_T7ktsZO62k7CaibgQvA

More from Aayush Ostwal and Towards Data Science





Aayush Ostwal in Towards Data Science

Test for existence of a Trend in a Time Series

Relative Order Test for testing the existence of a Trend in a Time series.

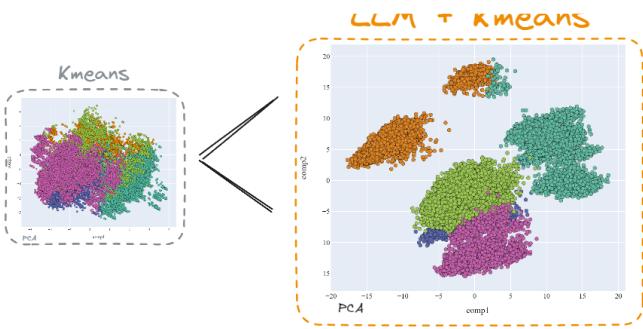
5 min read · Dec 22, 2020

👏 104

💬 3



...



Damian Gil in Towards Data Science

Mastering Customer Segmentation with LLM

Unlock advanced customer segmentation techniques using LLMs, and improve your...

24 min read · Sep 26

👏 3.6K

💬 26



...



Adrian H. Raudaschl in Towards Data Science

Forget RAG, the Future is RAG-Fusion

The Next Frontier of Search: Retrieval Augmented Generation meets Reciprocal...

⭐ · 10 min read · Oct 5

👏 2K

💬 24



...



Aayush Ostwal in Towards Data Science

Five Obstacles faced in Linear Regression

These five obstacles may occur when you train a linear regression model on your data...

7 min read · Jan 2, 2021

👏 27

💬



...

[See all from Aayush Ostwal](#)

[See all from Towards Data Science](#)

Recommended from Medium



 Ann Mary Shaju in Towards AI

Univariate, Bivariate, and Multivariate Analysis

A beginner guide to exploratory data analysis using Matplotlib and Seaborn

6 min read · May 18

 273  1

  ...



 Miriam Santos in Towards Data Science

A Data Scientist's Essential Guide to Exploratory Data Analysis

Best Practices, Techniques, and Tools to Fully Understand Your Data

11 min read · May 30

 1K  10

  ...

Lists



Predictive Modeling w/ Python

20 stories · 546 saves



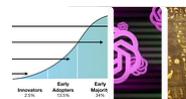
Practical Guides to Machine Learning

10 stories · 625 saves



The New Chatbots: ChatGPT, Bard, and Beyond

12 stories · 175 saves



Natural Language Processing

769 stories · 350 saves



Ebo Jackson in Level Up Coding

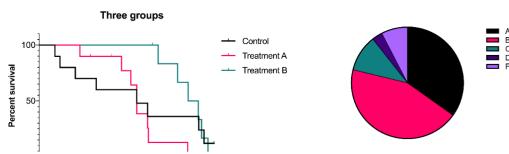
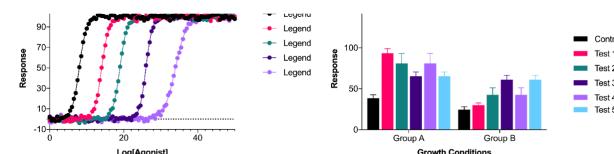
Explore The 10 Most Useful Charts in the 3 Most Popular Visualizatio...

Introduction

33 min read · Jul 7



...

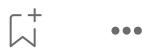


Carla Mota Leal

The Art of Data Visualization: Designing high impact and...

Data visualization is a powerful tool that enables us to unlock insights and...

5 min read · Jun 16



...



Mochamad Kautzar Ichramsyah in CodeX

Automate the exploratory data analysis (EDA) to understand the...

What is EDA?

11 min read · Jul 11



...

matplotlib

Aa U I

That's it ! Code Snippets

How to Change Font Style and Size in Matplotlib

Table of Contents

• 3 min read · May 8



...

[See more recommendations](#)