



Search



Write



R

An Ultimate Cheat Sheet for Data Visualization Techniques in Seaborn



Vildansarikaya · Follow

Published in Clarusway · 5 min read · Dec 26, 2020

848

9



Data visualization signs to the techniques used to information by encoding it as visual objects such as bars, points and lines in graphics. The goal is to communicate data and information to users in a clear and efficient manner. Data visualization is helpful to understand the structure of the data, determine the outlier, see the trends or patterns of the data, evaluate the results, and communicate these findings effectively to audience. This is one of the steps in data analysis, and this step is called with Exploratory Data Analysis (EDA).

Python provides many visualization packages for EDA. One of them is Seaborn library. Seaborn library makes statistical graphics in python, and it helps users explore and understand the data with graphics.

Here will be explained Seaborn plot types, how to plot with real data, when to use them, and finally provide a brand-new, broadly explained cheat sheet.

Let's start with importing required libraries and loading the dataset which contains some built-in datasets in seaborn. The “tips” dataset, one of built-in datasets, will be used.

```
import the libraries
In [1]: import numpy as np
        import pandas as pd
        import matplotlib.pyplot as plt
        import seaborn as sns

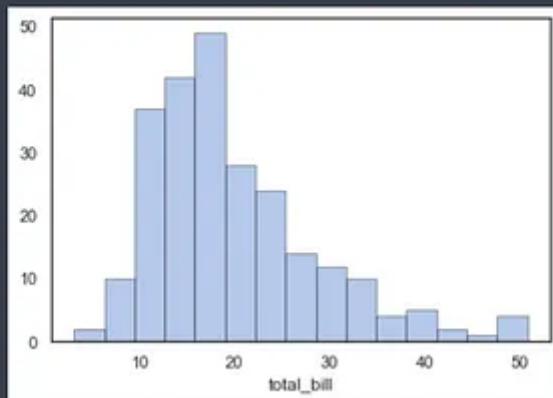
load the datasets
In [13]: tips = sns.load_dataset('tips')
          tips.head()

total_bill    tip    sex   smoker  day   time  size
0      16.99  1.01  Female     No    Sun  Dinner  2
1      10.34  1.66   Male     No    Sun  Dinner  3
2      21.01  3.50   Male     No    Sun  Dinner  3
3      23.68  3.31   Male     No    Sun  Dinner  2
4      24.59  3.61  Female     No    Sun  Dinner  4
```

1.DISTRIBUTION PLOT

A) **Distplot:** A distplot is used a univariate distribution of observation.

```
In [22]: sns.distplot(tips['total_bill'], hist_kws=dict(edgecolor="k", linewidth=1), bins=15, kde=False);
```



tips['total_bill']: It is a numeric column in tips dataset.

hist_kws: The keyword argument to change histogram format.

bins: It is used to set the number of bins, and it depends on your datasets.

kde: Kernel Density Estimate.

```
In [42]: sns.set(style="darkgrid")

# Set up the matplotlib figure
f, axes = plt.subplots(2, 2, figsize=(7, 7), sharex=True)
sns.despine()

# Generate a random univariate dataset
rs = np.random.RandomState(10)
d = rs.normal(size=100)

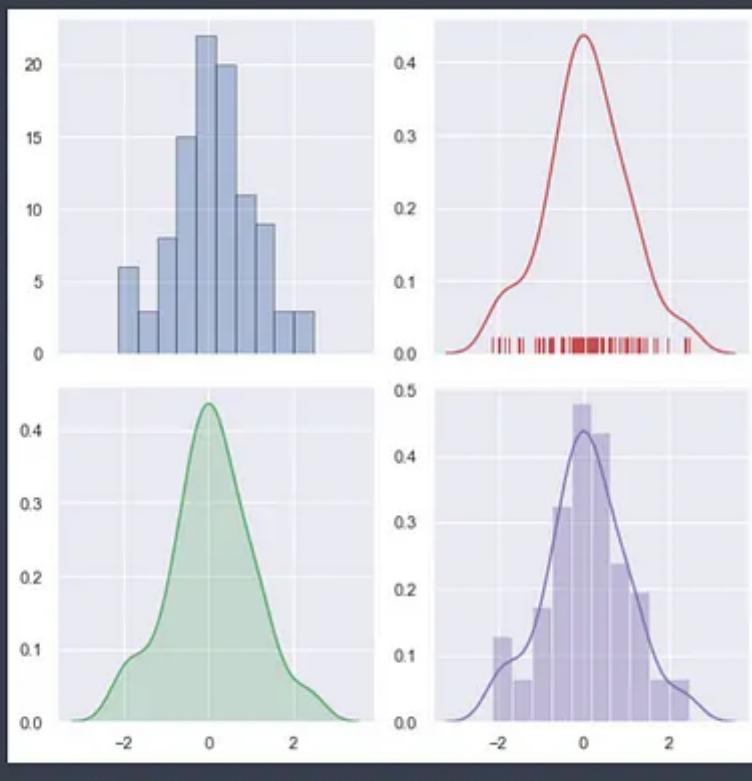
# Plot a simple histogram with binsize determined automatically
sns.distplot(d, kde=False, hist_kws=dict(edgecolor="k", linewidth=1), color="b", ax=axes[0, 0])

# Plot a kernel density estimate and rug plot
sns.distplot(d, hist=False, rug=True, color="r", ax=axes[0, 1])

# Plot a filled kernel density estimate
sns.distplot(d, hist=False, color="g", kde_kws={"shade": True}, ax=axes[1, 0])

# Plot a histogram and kernel density estimate
sns.distplot(d, color="m", ax=axes[1, 1])

plt.tight_layout()
```



2.CATEGORICAL PLOT

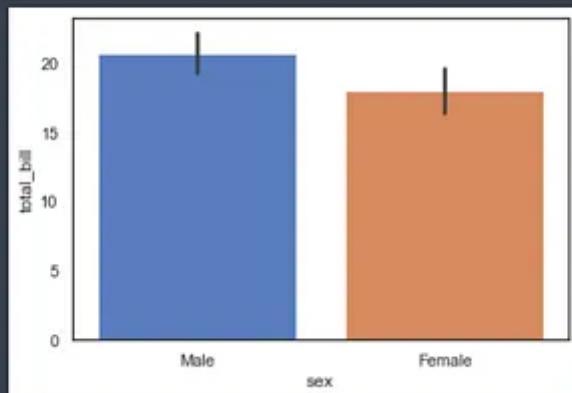
A categorical variable has two or more categories, nonnumeric values. For instance, “sex” column is a categorical variable column, having two categories (male and female) at “tips” dataset. “day” column is a categorical variable column, having a number of categories (thursday, friday, saturday, sunday, etc.) at “tips” dataset. Categorical plot can be used to visualize categorical variables.

A)Bar Plot: Bar plot is useful for displaying relationships between categorical variable and numerical variable. Bar plot has two methods. Dataset’s columns can be used or dataset’s columns assigned to x and y parameters are used.

Method 1

method 1

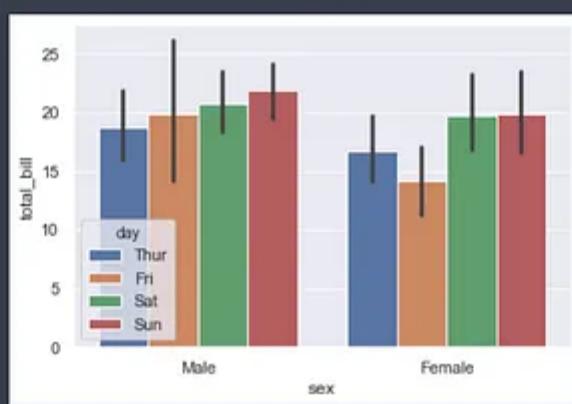
```
In [15]: sns.barplot(tips["sex"], tips["total_bill"]);
```



Method 2

method 2

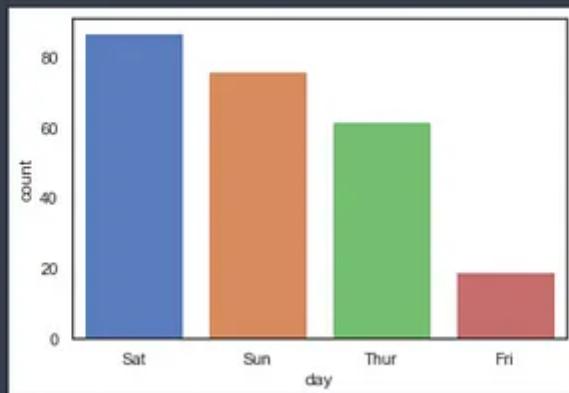
```
In [6]: sns.barplot(x="sex", y="total_bill", hue="day", data=tips);
```



hue: determines which column in the data frame should be used for colour encoding.

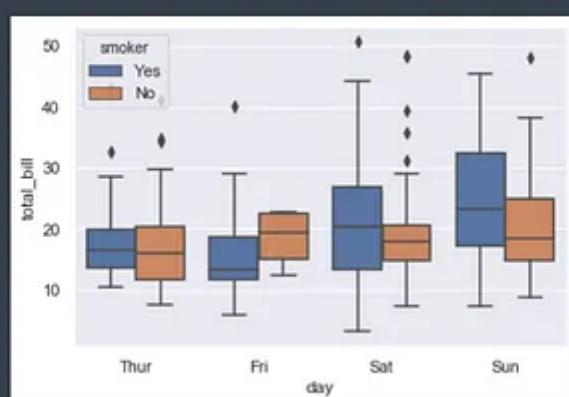
B) Count Plot: Count plot shows the counts of observations in each categorical variable.

```
In [27]: sns.countplot(x='day', order=tips.day.value_counts().index, data=tips);
```



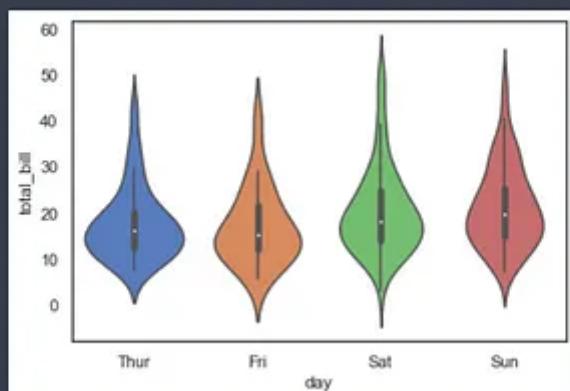
C)Box Plot: Box plot visualizes summarizing numeric data over a set of categorical variables, and it provides some information about data. This information consists of minimum score, maximum score, first quartile (25% of data), second quartile-median (50% of data), and third quartile (75% of data).

```
In [8]: sns.boxplot(x="day", y="total_bill", hue="smoker", data=tips);
```



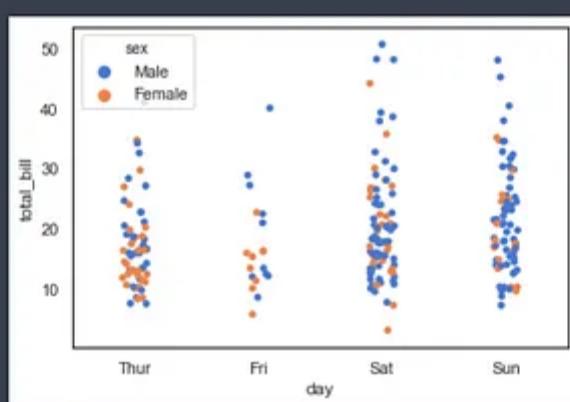
D)Violin Plot: Violin plot is a similar role as a box plot.

```
In [17]: sns.violinplot(x="day", y="total_bill", data=tips);
```



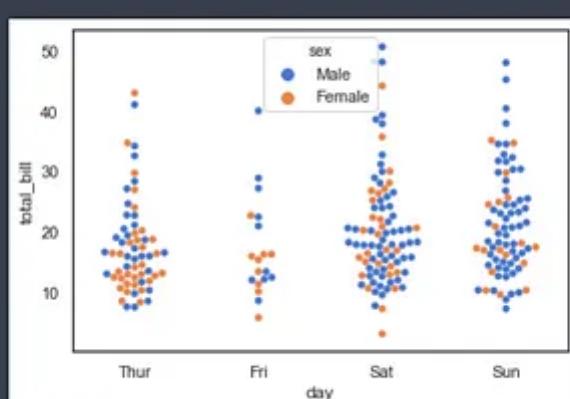
E) **Strip Plot:** Strip plot is a scatter plot. It shows the relationship between two variables.

```
In [18]: sns.stripplot(x="day", y="total_bill", hue="sex", data=tips);
```



F) **Swarm Plot(Violin Plot + Strip Plot):** As the name suggests, it consists of the combination of a violin plot and a strip plot.

```
In [19]: sns.swarmplot(x="day", y="total_bill", hue="sex", data=tips);
```



G)Cat Plot(Former Name: Factor Plot): This plot provides the relationship between a numerical and one or more categorical variables with several visual options.

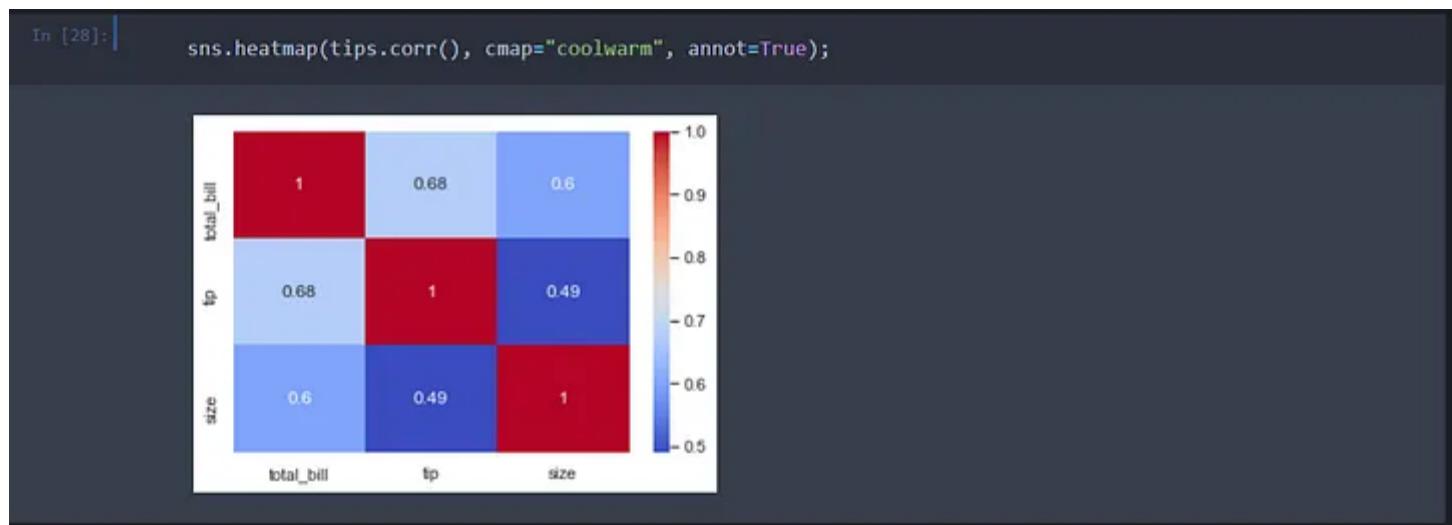


H)Point Plot: Point plot shows an estimate of central tendency(center of the data distribution) for a numerical variable.



3.MATRIX AND GRID PLOT

A) Heat Map: The heat map shows which variables are correlated to each other. It helps to understand correlation, and it provides a colorful visualization about correlation.



B) Pair Plot: The pair plot shows pairwise relationships in a dataset.



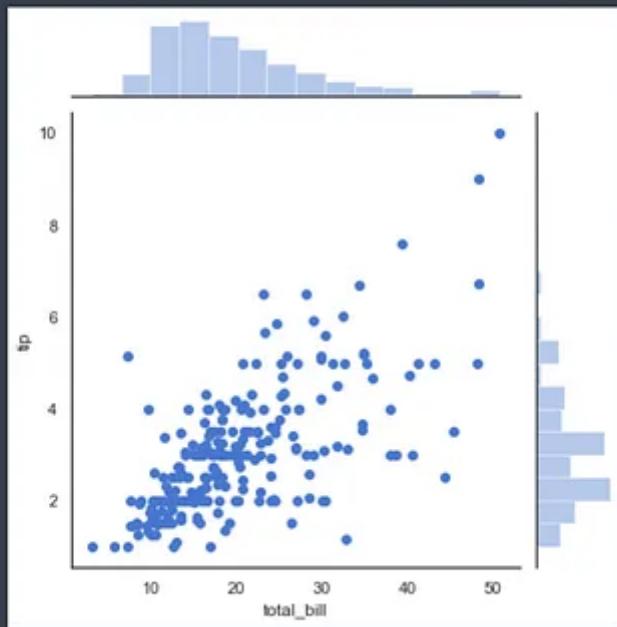
C)FacetGrid: FacetGrid helps to understand distribution of one variable as well as the relationship between multiple variables with using separate subsets.



4)JOINTPLOT

Jointplot visualizes two variables with bivariate and univariate graphs.

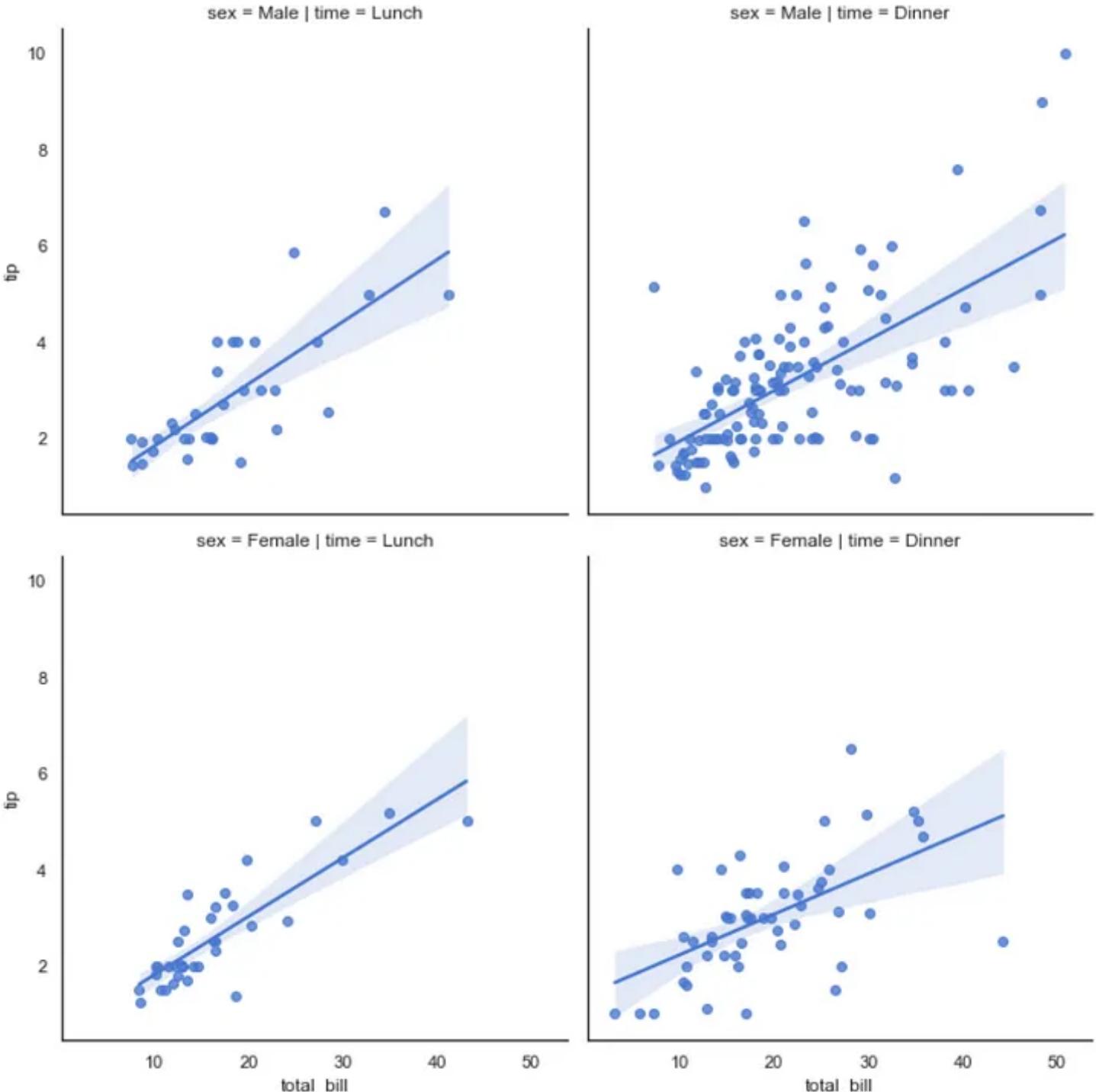
```
In [37]:  
sns.jointplot(x="total_bill", y="tip", kind="scatter", data=tips);  
# kind options =scatter (default), reg, resid, kde, hex.
```



5)LM PLOT

Lmplot can be used for drawing a scatter plot onto a FacetGrid.

```
In [19]:  
sns.lmplot(x="total_bill", ci=95, y="tip", data=tips, row="sex", col="time", aspect=1, height=5)
```



To sum up:

- Seaborn is a Python data visualization library built on Matplotlib.
- It provides a high-level interface for drawing attractive and informative statistical graphics.
- Seaborn is helpful to understand structure of the data, determine the outliers, and understand many things through visualization.

- Seaborn has powerful data visualization capabilities for EDA. All charts mentioned in the article can be used for EDA.
- This cheat sheet is a useful reference where you can find all the examples in the article and easily access them whenever you want.

You can download the **cheat sheet** containing all the codes from [here](#).

Cheatsheet

Data Visualization

Seaborn

Data Science

Data Analysis



Written by Vildansarikaya

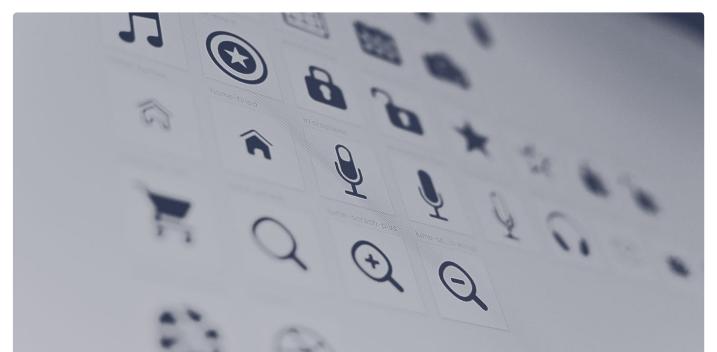
66 Followers · Writer for Clarusway

Data Scientist, Industrial Engineer

Follow



More from Vildansarikaya and Clarusway



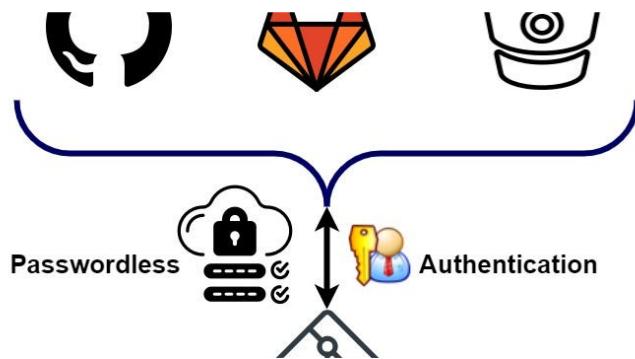
TYPES OF RELATIONSHIPS IN DATA MODELLING

Start by learning the meaning of the relational database. A relational database is a type of...

4 min read · Nov 24, 2020

 545  8

+ 



 Joe Blue in Clarusway

How to Use Git/GitHub without asking for authentication always:...

The authentication to access the source code repositories like GitHub, GitLab, BitBucket,...

7 min read · Dec 30, 2020

 944  9

+ 

[See all from Vildansarikaya](#)

[See all from Clarusway](#)

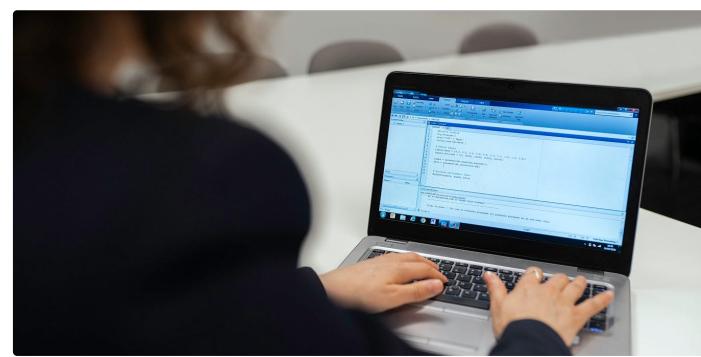
Setting up React Native Vector Icons for IOS

Icons tell more than a plenty of words do. This is one of the reasons that icons are widely...

4 min read · Dec 20, 2020

 1.3K  9

+ 



 Alex Duncan in Clarusway

Installing and Upgrading Ansible with *pip

You can install Ansible on many systems with pip. Pip is the main Python package manager.

3 min read · Jul 4, 2021

 435 

+ 



Mutiara Tambunan

Automated EDA and Creating a Simple Interactive Dashboard in...

In today's data-driven world, the ability to analyze and visualize data is fundamental for...

6 min read · Sep 10

263

...

Robert Daly in Python in Plain English

Preprocessing Data for Linear Regwith Scikit-Learn

This week I learned how to create a Linear Regression model. I think it's so cool that...

5 min read · Aug 15

20

...

Lists



Predictive Modeling w/ Python

20 stories · 659 saves



Practical Guides to Machine Learning

10 stories · 739 saves



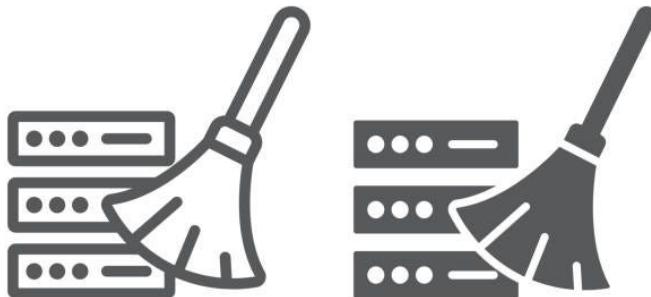
New_Reading_List

174 stories · 209 saves



ChatGPT prompts

30 stories · 747 saves



Nikhil Malkari

Manoj Das

Mastering Data Cleaning Techniques: Handling Missing...

INTRODUCTION

9 min read · Jun 21



2



1



...



Mochamad Kautzar Ichramsyah in CodeX

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

What is EDA?

11 min read · Jul 12



1.97K



23



...

Discover DataPrep: Make Exploratory Data Analysis Easier i...

What is DataPrep package in Python? How to do EDA using DataPrep? Example of EDA wit...

7 min read · Jul 7



3



1



...



Bradley Stephen Shaw in Towards Data Science

Make Your Tables Look Glorious

Simple formatting tricks to make your pandas DataFrames presentation-ready

13 min read · Jan 11



440



8



...

[See more recommendations](#)