Workshop on Python (Day 5)

By Suriya G
Organized by Suresh Sir, UPNM

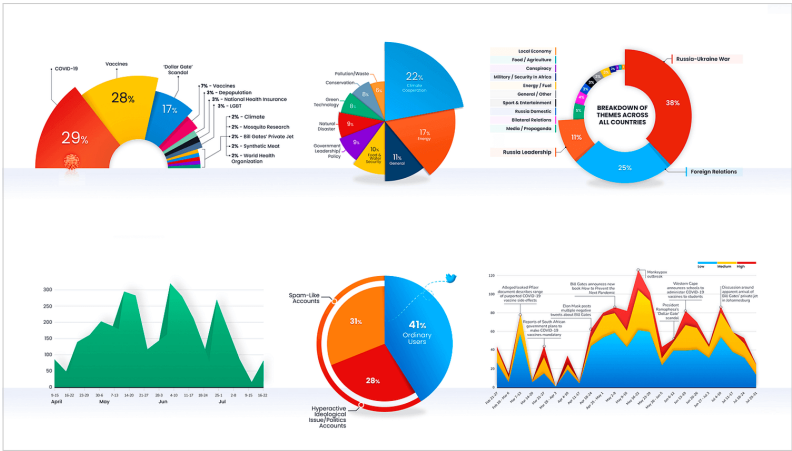
TABLE OF CONTENTS

What is Data Visualization?

Process of describing information through **Visual Rendering**

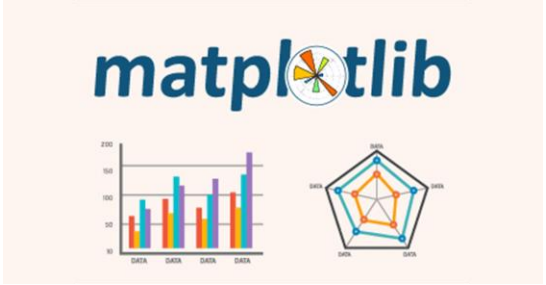
	IID - Interviewer Identification	Does respondent have a mobile phone?	Work status	Occupation	Age	Top of mind awareness
1	853.0	Yes	Fulltime worker	Associate professional clerical, sales and services	45-54 yrs	Optus
2	854.0	Yes	Fulltime worker		20-24 yrs	Optus
3	855.0	Yes	Retired		45-54 yrs	Optus
4	851.0	Yes	Fulltime worker	manager/ administrator	25-29 yrs	Optus
5	852.0	Yes	Student		20-24 yrs	Optus
6	883.0	Yes	Fulltime worker	Associate professional clerical, sales and services	45-54 yrs	Telstra (Mobile Net)
7	884.0	Yes	Fulltime worker		20-24 yrs	Vodafone
8	885.0	Yes	Retired		45-54 yrs	Optus
9	881.0	Yes	Student		16-19 yrs	Optus
10	882.0	Yes	Fulltime worker	manager/ administrator	20-24 yrs	Telstra (Mobile Net)
11	713.0	Yes	Part-time worker		20-24 yrs	Optus

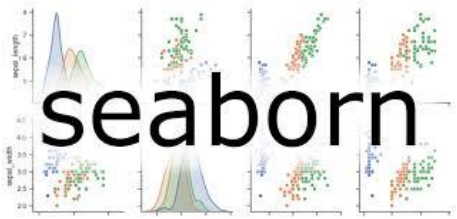
Showing 1 to 11 of 725 rows.



Visualization libraries

- matplotlib
- Seaborn
- Bokeh
- Plotly etc….



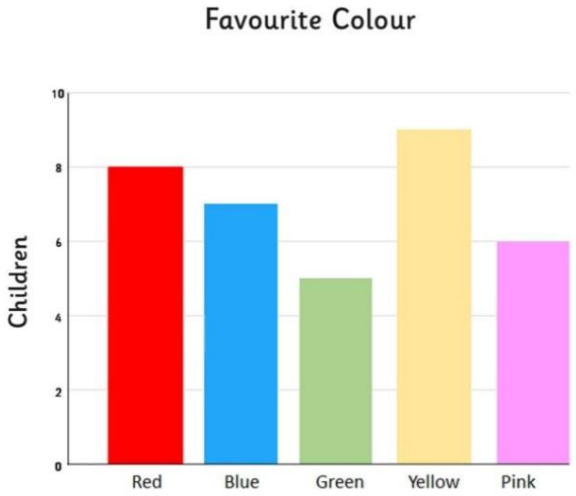


bokeh

plotly

Bar Graph

- **Category comparison**
- Used for comparing different categories or groups
- Example: Sales analysis across regions

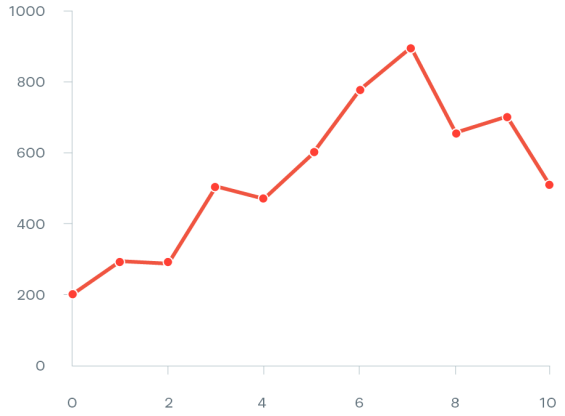


bar.py

```
1  # Basic syntax for bar chart
2  plt.bar()                    #Method 1
3  sns.countplot(x,data)        #Method 2
```

Line Graph

- **Trend over time**
- Shows changes over time, useful for tracking trends
- Example: Sales trend over months

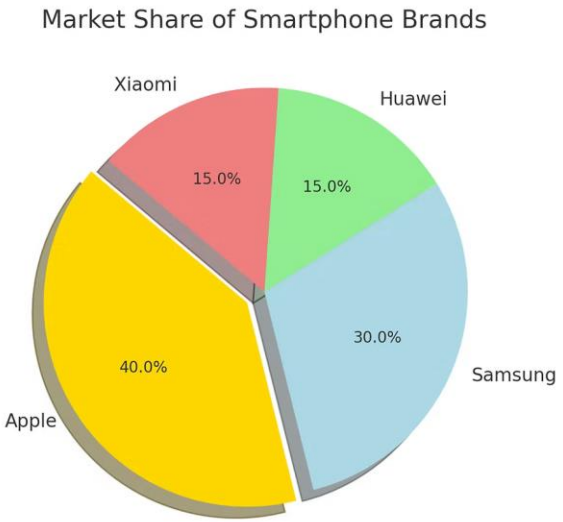


line.py

```
1  # Basic syntax for line chart
2  plt.plot(x,y)                    #Method 1
3  sns.lineplot(x,y,data)           #Method 2
```

Pie Chart

- **Proportion distribution**
- Displays percentage or proportional breakdown of a whole
- **Example: Sales across each category**



pie.py

```
# Basic syntax for pie chart
plt.pie(values,labels,autopct)
```

Scatter Plot

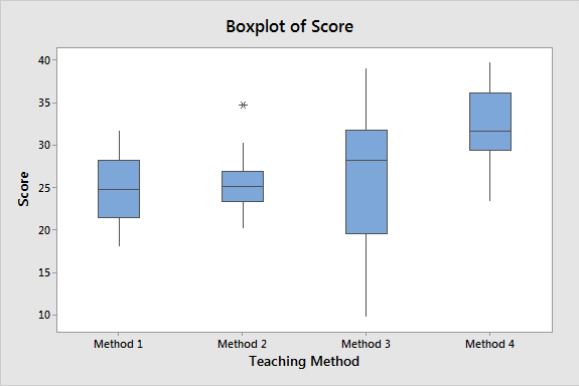
- **Correlation analysis**
- Helps identify relationships between two numerical variables
- **Example: Sales vs Profit distribution**

scatter.py

```
# Basic syntax for scatter plot
plt.scatter(x,y)                    #Method 1
sns.scatterplot(x,y,data)           #Method 2
```

Box Plot

- Summarizes distribution **(median, quartiles, outliers)** of numerical data.
- Useful for comparing distributions across categories.

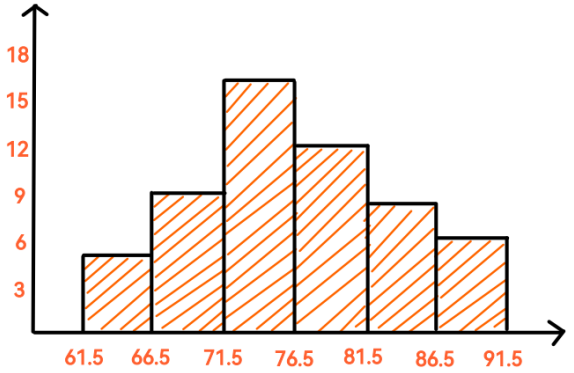


box.py

```
1  # Basic syntax for boxplot
2  df.plot.box()
```

Histogram

- Displays the distribution of a single numerical variable.
- Bars represent frequency or count within bins.



hist.py

```
1  # Basic syntax for histogram
2  df.plot.hist(bins=5)
```

How has the average salary changed over the years?

Graph Type: Line Chart

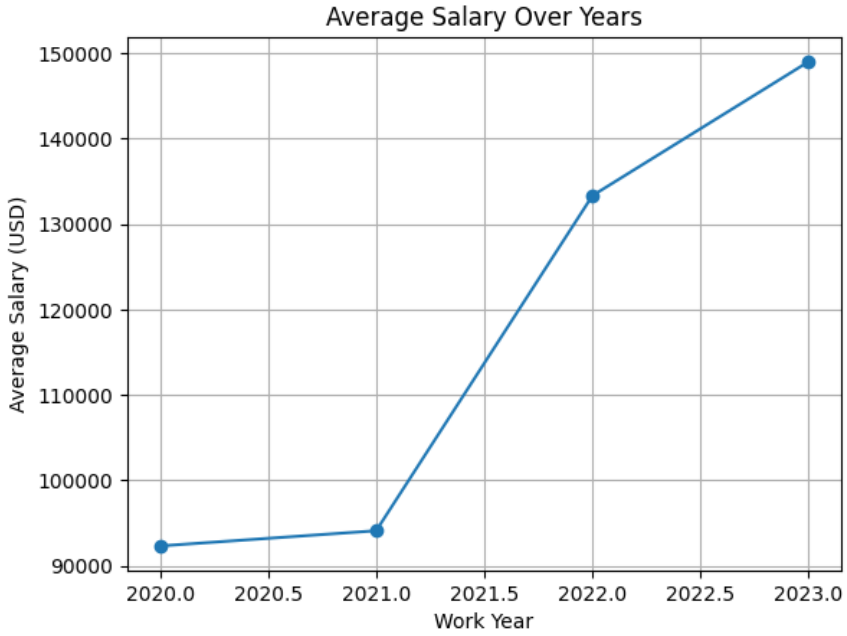
How has the average salary changed over the years?

Graph Type: Line Chart

Line_Chart.py

```
# Basic syntax for histogram
df.groupby("work_year")["salary_in_usd"].mean().plot
(kind="line", marker='o', title="Average Salary Over
Years")
plt.xlabel("Work Year")
plt.ylabel("Average Salary (USD)")
plt.grid(True)
plt.show()
```

How has the average salary changed over the years?



What is the distribution of employee experience levels?

Graph Type: Count Plot (Bar)

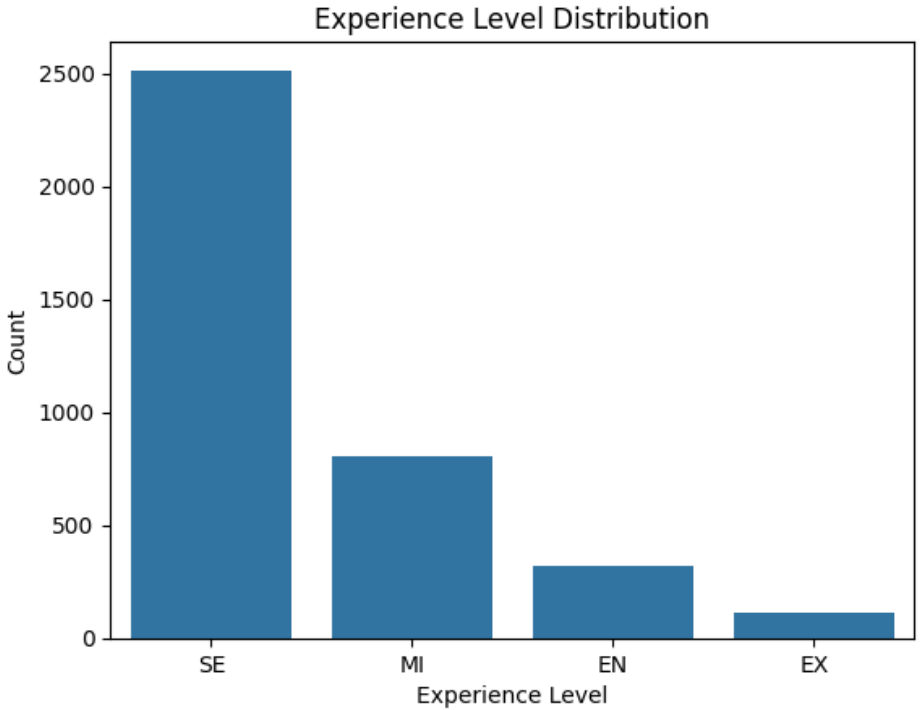
What is the distribution of employee experience levels?

Graph Type: Count Plot (Bar)

bar.py

```
1  import seaborn as sns
2  sns.countplot(data=df, x="experience_level")
3  plt.title("Experience Level Distribution")
4  plt.xlabel("Experience Level")
5  plt.ylabel("Count")
6  plt.show()
```

What is the distribution of employee experience levels?



Which job titles pay the most on average?

Graph Type: Bar Chart

Which job titles pay the most on average?

Graph Type: Bar Chart

hist.py

```
1  # Basic syntax for histogram
2  top_jobs =
   df.groupby("job_title")["salary_in_usd"].mean().sort_values().head(10)
3  top_jobs.plot(kind="barh", title="Top 10 Highest Paying Job Titles")
4  plt.xlabel("Average Salary (USD)")
5  plt.gca().invert_yaxis()
6  plt.show()
```

Which job titles pay the most on average?

What is the relationship between remote ratio and salary?

Graph Type: Scatter Plot

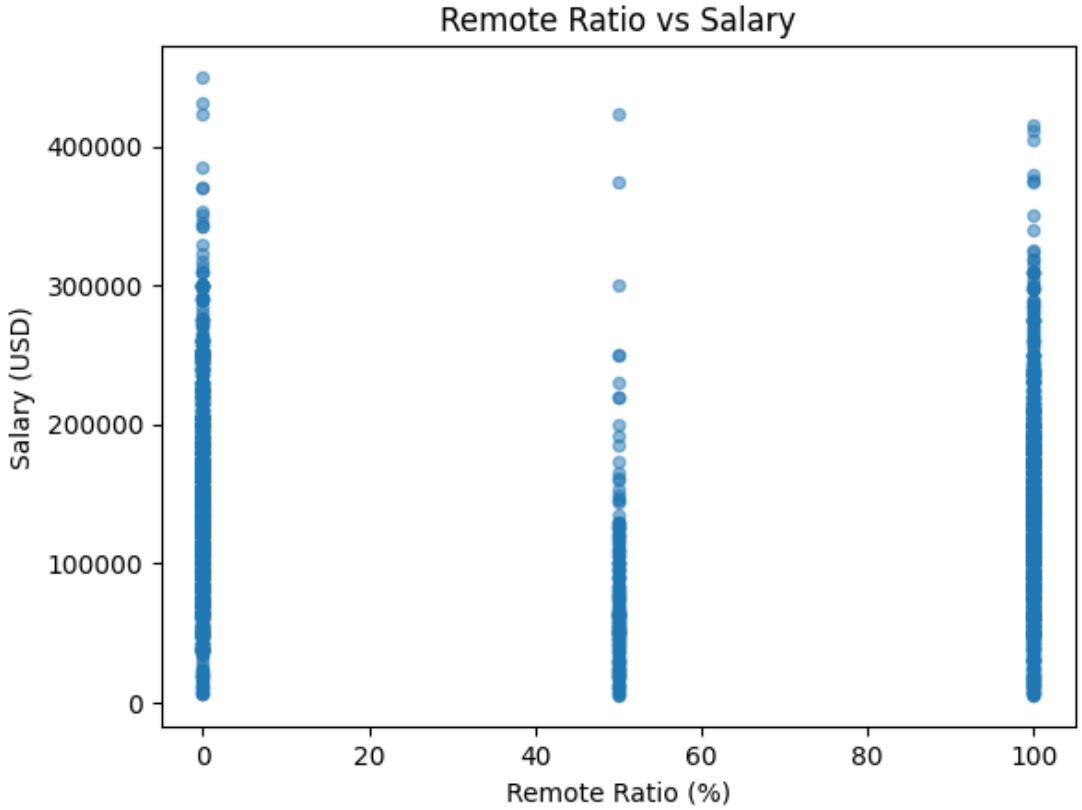
What is the relationship between remote ratio and salary?

Graph Type: Scatter Plot

hist.py

```
1  # Basic syntax for histogram
2  df.plot.scatter(x="remote_ratio", y="salary_in_usd", title="Remote
3  Ratio vs Salary")
4  plt.xlabel("Remote Ratio (%)")
5  plt.ylabel("Salary (USD)")
6  plt.show()
```

Which job titles pay the most on average?



How many employees work under each employment type?

Graph Type: Pie Chart

How many employees work under each employment type?

Graph Type: Scatter Plot

hist.py

```
1  # Basic syntax for histogram
2  df["employment_type"].value_counts().plot.pie(autopct="%1.1f%%",
3  title="Employment Type Distribution")
4  plt.ylabel("")
5  plt.show()
```

How many employees work under each employment type?