Advanced Visualization tools

Data Science with Python

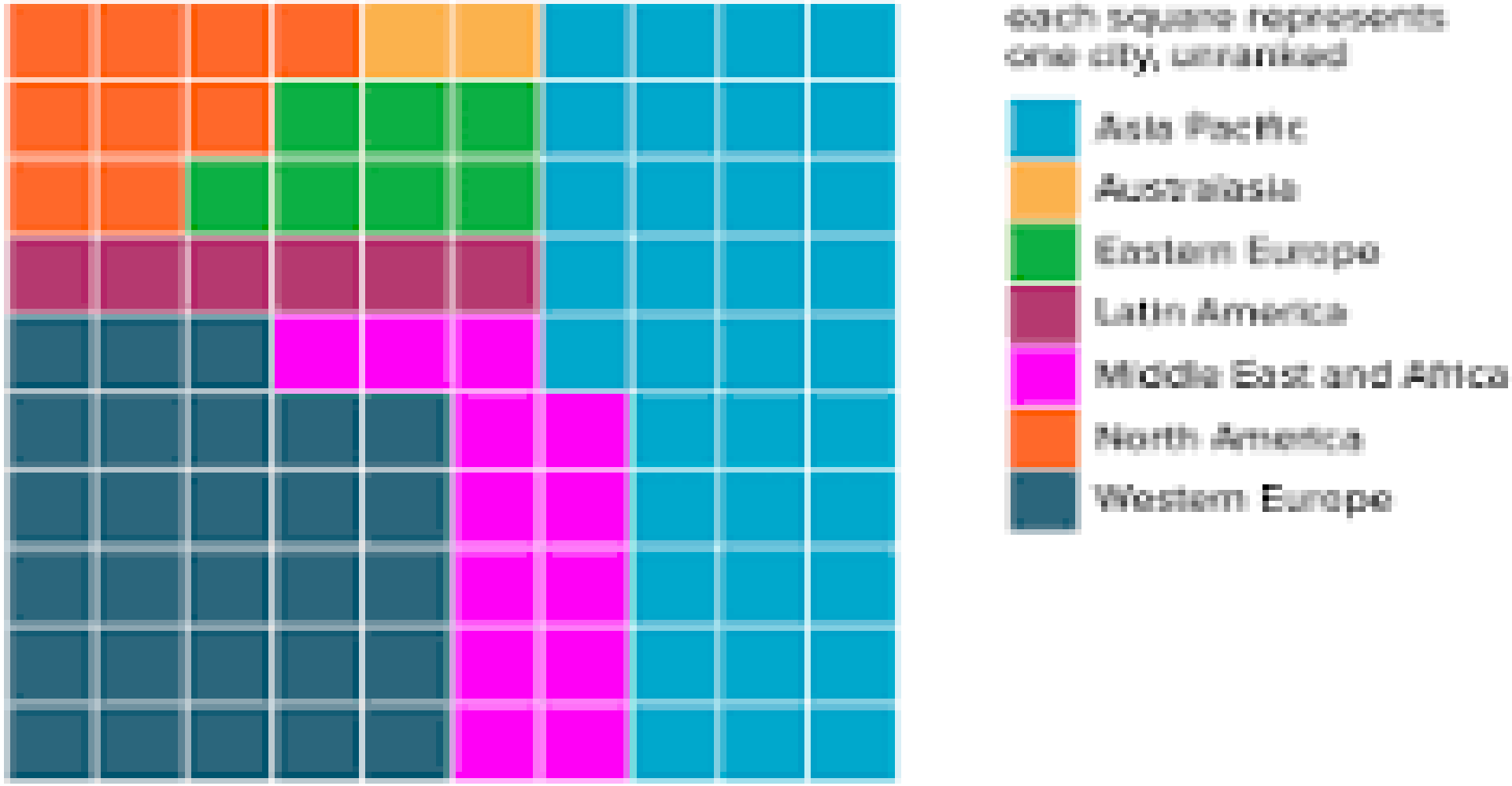
K.Krishna priya

Cse department

Waffle chart using pyWaffle in Python

- A Waffle Chart is a gripping visualization technique that is normally created to display progress towards goals.
- Where each cell in the Waffle Chart constitutes of 10 X 10 cell grid in which each cell represents one percentage point summing up to total 100%.
- It is commonly an effective option when you are trying to add interesting visualization features to a visual. Waffle Charts are widely used as an Excel dashboard.
- For generating Waffle Chart in Python, modules needed are – matplotlib, pandas and pyWaffle.
- To install these packages, run the following commands :
- pip install matplotlib
- pip install pandas
- pip install pywaffle

Where Are the Top 100 City Destinations?

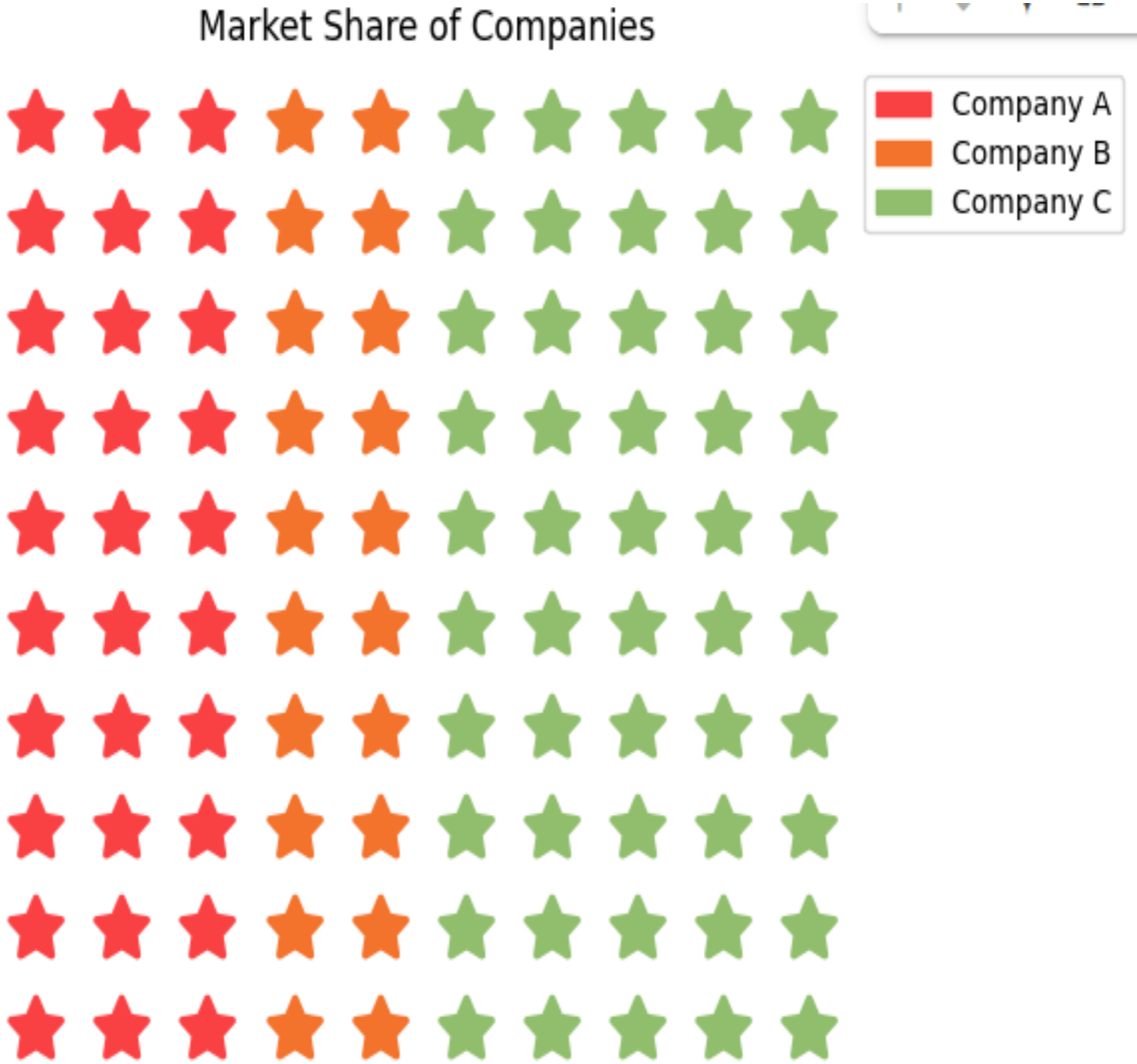


When to Use Waffle Charts

- Waffle charts are ideal for visualizing **part-to-whole relationships** (similar to pie charts).
- Use it when you want to represent **percentage-based data** or categories.
- Suitable for survey results, market share, or demographic data.
- The chart helps compare the contribution of each company in a single view.
- Easier to interpret when percentages are essential.

Example Data

- Market share of companies:
- data = {'Company A': 30, 'Company B': 20, 'Company C': 50}
- **import matplotlib.pyplot as plt**
- **from pywaffle import Waffle**
- **# Data representing market share**
- **data = {'Company A': 30, 'Company B': 20, 'Company C': 50}**
- **# Plotting the waffle chart**
- **plt.figure(FigureClass=Waffle, rows=10, values=data, colors=["#f94144", "#f3722c", "#90be6d"], legend={'loc': 'upper left', 'bbox_to_anchor': (1, 1)}, icons='star', icon_size=20)**
- **plt.title('Market Share of Companies')**
- **plt.show()**

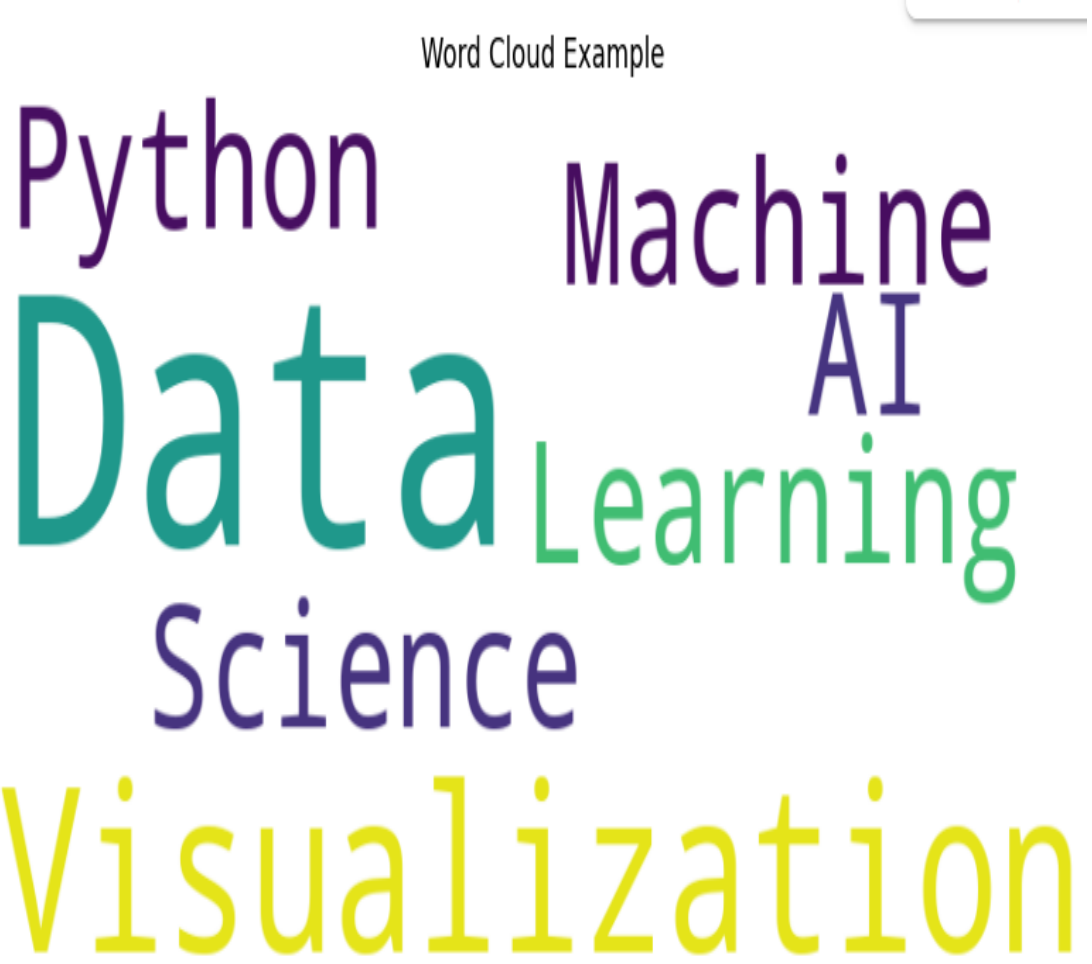


2. When to Use Word Clouds

- Word clouds are perfect for visualizing text data.
- Use them to represent the frequency of words in articles, social media posts, surveys, or customer reviews.
- Larger words indicate higher frequency.
- Example Data
- Sample text data:
- text = "Python Data Visualization Machine Learning Data Science AI Visualization"

Example

- from wordcloud import WordCloud
- import matplotlib.pyplot as plt
- # Creating a WordCloud
- wordcloud = WordCloud(width=800, height=400, background_color='white').generate(text)
- # Displaying the WordCloud
- plt.figure(figsize=(10, 5))
- plt.imshow(wordcloud, interpolation='bilinear')
- plt.axis("off")
- plt.title("Word Cloud Example")
- plt.show()



Regression Plots using Seaborn

- Use regression plots to visualize the relationship between two continuous variables.
- Suitable for trend analysis, correlation studies, and predictive modeling.
- Example Data
- tips dataset (from Seaborn): Contains restaurant bills and tips.

Example program

- import seaborn as sns
- import matplotlib.pyplot as plt
-

  # Load example data
- tips = sns.load_dataset("tips")
-

  # Create a regression plot
- sns.regplot(x="total_bill", y="tip", data=
- plt.title('Regression Plot: Bill vs Tip')
- plt.show()

4. Map with Marker using Folium

- Plotting geographical points like city locations, business branches, or events.
- Example Location:

- Bangalore (Latitude: 12.9716, Longitude: 77.5946)

- Expected Map Output:

- An interactive map with a marker labeled "Bangalore".
- **You can zoom in and out** for more detail using Folium’s map controls.

Example

- import folium
- # Create a map centered at Bangalore
- mymap = folium.Map(location=[12.9716, 77.5946], zoom_start=10)
- # Add a marker for Bangalore
- folium.Marker([12.9716, 77.5946], popup="Bangalore", tooltip="Bangalore City").add_to(mymap)
- # Save the map
- mymap.save("map_with_marker.html")
- print("Map saved as map_with_marker.html")

5. Choropleth Map

- Visualizing regional data like population, GDP, or election results.
- Choropleth maps are useful for visualizing data distributed over geographical areas.
- Suitable for population data, election results, or economic statistics.
- The color gradient visually highlights areas with higher or lower population density.
- Helpful for decision-making in fields like urban planning, sales, and public policy.

Example program

- state_data = pd.DataFrame({'State': ['California', 'Texas', 'Florida'], 'Population': [39500000, 29000000, 22000000]}) import foliumimport pandas as pd
- # Example state
- datastate_data = pd.DataFrame({ 'State': ['California', 'Texas', 'Florida'], 'Population': [39500000, 29000000, 22000000]})
- # GeoJSON URL for US state boundaries
- geojson_url =“ https://raw.githubusercontent.com/python-visualization/folium/master/examples/data/us-states.json”
- # Create mapm = folium.Map(location=[37.0902, -95.7129], zoom_start=4)
- # Add Choropleth layer
- folium.Choropleth(geo_data=geojson_url, name='choropleth', data=state_data, columns=['State', 'Population'], key_on='feature.id', fill_color='YlOrRd', fill_opacity=0.7, line_opacity=0.2, legend_name='Population in USA').add_to(m)
- # Save map
- m.save("choropleth_map.html")
- print("Choropleth Map saved as choropleth_map.html")

When to Use Each Visualization

Visualization	Best For	Example Data
Waffle Chart	Part-to-whole comparisons	Market Share, Survey Results
Word Cloud	Text data exploration	Customer Reviews, Survey Feedback
Regression Plot	Continuous variable relationship analysis	Sales Data, Financial Data
Folium Marker Map	Geospatial data with specific locations	Store Locations, Landmark Data
Choropleth Map	Regional data comparisons	Population, Election Results