Bar Charts & Plotting Pivot Tables

1

Bar Charts

- Bar charts are very similar to histograms which are used to represent the distribution of values in data
- While a histogram typically represent the *frequency distribution of continuous variables*, a bar chart is a *comparison of discrete variables*
- Another way to think about it is, a histogram *presents numerical data* and a bar chart *shows categorical data*

2

Creating a Bar Chart Using the *bar* Method

- One way to create a bar chart is by using matplotlib's *bar* method
- This is similar to the *hist* method, where you provide the data to plot and some additional parameter configurations

```
import pandas as pd
import matplotlib.pyplot as plt

plt.bar(
    data_for_x_axis,
    data_for_y_axis,
    #other optional parameters ...
)
```


3

Creating a Bar Chart Using the *plot* Method

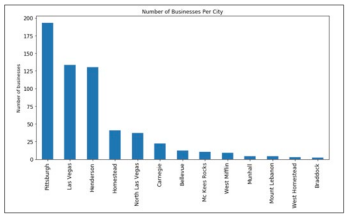
- Another *very easy* way to create a bar chart, is by using a DataFrame's built-in *plot* method
- Remember when we were working with the Yelp data, and we asked the question "How many businesses are there in each *city*?" Here's an example where we visualize that data:

```
#count the records for each city and get a new DataFrame
df_city_value_counts = df['city'].value_counts()

#call the plot method and set the kind parameter to 'bar'
df_city_value_counts.plot(kind='bar', figsize=(12, 6), fontsize=12,
   legend=False, title="Number of Businesses Per City")

plt.ylabel("Number of businesses")
plt.show()
```


4

Creating a Bar Chart Using the *plot* Method

- Another *very easy* way to create a bar chart, is by using a DataFrame's built-in *plot* method
- Remember when we were working with the Yelp data, and we asked the question "How many businesses are there in each *city*?" Here's an example where we visualize that data:



5

Plotting a Pivot Table

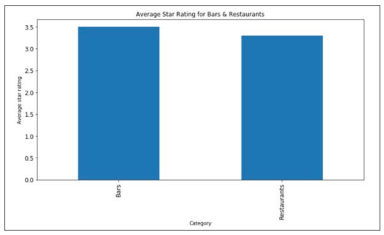
- Here's a visualization of a pivot table that displays the average (mean) star rating for bars and restaurants:

```
bar_rest = df["category_0"].isin(["Bars", "Restaurants"])
df_bar_rest = df[bar_rest]

#pivot along category
pivot_state_cat = pd.pivot_table(df_bar_rest, index=["category_0"])

#filter the df_bar_rest DataFrame columns
pivot_state_cat = pivot_state_cat[["stars"]]

#call the plot method and set the kind parameter to 'bar'
pivot_state_cat.plot(kind='bar', figsize=(12, 6), fontsize=12,
    legend=False, title="Average Star Rating for Bars & Restaurants")

plt.xlabel("Category")
plt.ylabel("Average star rating")
plt.show()
```


6

Plotting a Pivot Table

- Here's a visualization of a pivot table that displays the average (mean) star rating for bars and restaurants:



Penn Engineering

7