**Dipika Jothinathan – Pandas Homework**

import pandas as pd  
import matplotlib.pyplot as plt  
  
student\_data = {  
 ‘name’: [‘Paul’, ‘Tom’, ‘Tim’, ‘Kathy’, ‘Anastasia’, ‘Emily’, ‘Robert’, ‘Bob’, ‘Laura’, ‘Michael’],  
 ‘marks’: [68, 78, 98, 74, 67, 89, 81, 76, 91, 86],  
 ‘attempts’: [1, 3, 2, 1, 1, 2, 3, 3, 1, 2],  
 ‘label’: [‘a’, ‘b’, ‘c’, ‘d’, ‘e’, ‘f’, ‘g’, ‘h’, ‘I’, ‘j’]  
}  
  
print(“\n1. Using the dictionary, create a data frame.”)  
datframe = pd.DataFrame(student\_data)  
print(datframe, “\n”)  
  
print(“2. Change the indices to use labels column.”)  
df = datframe.set\_index(‘label’)  
print(df, “\n”)  
  
print(“3. Create and display summary for each column.”)  
summary = df.describe()  
print(summary, “\n”)  
  
print(“4. Create and display a new data frame by extracting first 3 rows and 3 columns.”)  
df1 = df[0:3]  
print(df1, ‘\n’)  
  
print(“5. Create and display a new data frame by extracting name & marks to include only rows where attempts <=2.”)  
df2 = df[df.attempts <= 2]  
df2 = df2[[‘name’, ‘marks’]]  
print(df2, ‘\n’)  
  
print(“6. Create and display a new data frame by extracting name & marks to include only rows where attempts <=2 and score >80.”)  
df3 = df[(df.attempts <= 2) & (df.marks > 80)]  
df3 = df3[[‘name’, ‘marks’]]  
print(df3, ‘\n’)  
  
print(“7. Add a new new row ‘k’ to the data frame – ‘name’: ‘Tracy’, ‘marks’: 82, ‘attempts’: 1, ‘label’: ‘k’.”)  
row\_data = {‘name’: ‘Tracy’, ‘marks’: 82, ‘attempts’: 1, ‘label’: ‘k’}  
df\_row = datframe.append(row\_data, ignore\_index= True)  
df\_row = df\_row.set\_index(‘label’)  
print(df\_row, ‘\n’)  
  
print(“8. Add a new column ‘gender’ with values [‘M’,’M’,’M’,’F’,’F’,’F’,’M’,’F’,’M’,’F’].”)  
df[‘gender’] = [‘M’, ‘M’, ‘M’, ‘F’, ‘F’, ‘F’, ‘M’, ‘F’, ‘M’, ‘F’]  
print(df, ‘\n’)  
  
print(“9. Sort the values in order of marks column in ascending order and display data.”)  
df\_sort = df.sort\_values(‘marks’)  
print(df\_sort, ‘\n’)  
  
print(“10. Create a duplicate column ‘c\_marks’ by subtracting 2 from marks for each student.”)  
df[‘c\_marks’] = df.marks – 2  
print(df, ‘\n’)  
  
print(“11. Delete marks column.”)  
df\_drop = df.drop([‘marks’], axis = 1)  
print(df\_drop, ‘\n’)  
  
print(“12. Plot a bar chart that displays names of students on x axis and marks on y axis.”)  
df.plot.bar(x=’name’, y=’marks’, legend = False, color=list(‘rgbkymc’))  
plt.xlabel(‘Names’)  
plt.ylabel(‘Marks’)  
plt.title(“Student marks”)  
plt.xticks(rotation=90)  
plt.show()  
  
print(“13. Create a pie chart to represent % distribution of gender.”)  
df\_pie = df.groupby([‘gender’]).count()/len(df)\*100  
df\_pie.plot(kind=’pie’,y=’name’)  
plt.ylabel(‘Gender’)  
plt.title(“Pie chart to represent % distribution of gender.”)  
plt.show()  
  
print(“14. Create a histogram to show count of students in buckets of marks – 60-70, 70-80, 80-90, 90-100.”)  
df[‘marks’].plot.hist(bins=[60, 70, 80, 90, 100])  
plt.title(“Histogram to show count of students in marks buckets”)  
plt.xlabel(‘Marks’)  
plt.ylabel(‘Count’)  
plt.show()

A screenshot of a computer

Description automatically generated

A paper with text on it

Description automatically generated

A graph of different colored rectangular shapes

Description automatically generated

A pie chart with text and blue and orange circles

Description automatically generated A graph with blue squares

Description automatically generated