Assignment 9 - Advanced visualization

<u>**INSTRUCTIONAL DETAILS**</u>

Here are a few examples of data visualizations that you can create using Python:

Line chart: You can use a line chart to visualize trends over time or to compare multiple datasets.

```python
import matplotlib.pyplot as plt

# data
x = [1, 2, 3, 4, 5]
y1 = [1, 2, 3, 4, 5]
y2 = [2, 4, 6, 8, 10]

# create the plot
plt.plot(x, y1, label='line 1')
plt.plot(x, y2, label='line 2')

# add labels and title
plt.xlabel('X-axis')
plt.ylabel('Y-axis')
plt.title('Line chart example')

# add legend
plt.legend()

# show the plot
plt.show()
```

Bar chart: You can use a bar chart to compare values across categories.

```python
import matplotlib.pyplot as plt

# data
categories = ['Category 1', 'Category 2', 'Category 3']
values = [10, 20, 30]

# create the plot
plt.bar(categories, values)

# add labels and title
plt.xlabel('Categories')
plt.ylabel('Values')
plt.title('Bar chart example')

# show the plot
plt.show()
```

Scatter plot: You can use a scatter plot to visualize the relationship between two variables.

```python
import matplotlib.pyplot as plt

# data
```

```
x = [1, 2, 3, 4, 5]
y = [1, 4, 9, 16, 25]

# create the plot
plt.scatter(x, y)

# add labels and title
plt.xlabel('X-axis')
plt.ylabel('Y-axis')
plt.title('Scatter plot example')

# show the plot
plt.show()
```

Heatmap: You can use a heatmap to visualize the distribution of values in a matrix.

```
import seaborn as sns
import numpy as np

# data
data = np.random.rand(10, 12)

# create the plot
sns.heatmap(data)

# add title
plt.title('Heatmap example')

# show the plot
plt.show()
```

Box plot: You can use a box plot to visualize the distribution of values in a dataset and to identify outliers.

```
import seaborn as sns

# data
data = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]

# create the plot
sns.boxplot(data)

# add title
plt.title('Box plot example')

# show the plot
plt.show()
```

BACKGROUND

Advanced data visualizations are important to businesses because they allow them to effectively communicate and interpret complex data in a way that is easy for people to understand. By using advanced visualization techniques, businesses can more clearly communicate the insights and findings from their data analysis, which can help inform decision-making and drive business success.

Advanced data visualizations can also help businesses identify trends and patterns in their data that might not be immediately apparent using traditional visualization techniques. For example, using techniques such as multivariate visualization or interactive visualizations can help businesses understand how multiple variables relate to each other and how they change over time. This can help businesses identify opportunities for growth and improvement.

In addition, advanced data visualizations can help businesses engage and communicate with stakeholders more effectively. By presenting data in a visually appealing and easily understandable way, businesses can more effectively convey their message and persuade others to take action.

Overall, advanced data visualizations are important to businesses because they allow them to effectively communicate and interpret complex data, identify trends and patterns, and engage and communicate with stakeholders more effectively. This can ultimately help businesses make better decisions and achieve their goals.

RESEARCH QUESTION

Data analysis can be a valuable tool for Amtrak, the national passenger rail service in the United States, in a number of ways. Here are a few examples of how data analysis can be used to improve the operations and performance of Amtrak:

Route optimization: Data analysis can be used to analyze data on factors such as passenger demand, travel patterns, and the cost of operating different routes. This can help Amtrak optimize its route network and make more informed decisions about which routes to add or retire.

Capacity planning: Data analysis can be used to analyze data on factors such as passenger demand, travel patterns, and the availability of rolling stock (e.g. trains, cars, and locomotives). This can help Amtrak plan its capacity more effectively and ensure that it has the right amount of capacity to meet demand.

Asset management: Data analysis can be used to analyze data on the performance and maintenance of Amtrak's assets, such as its rolling stock and infrastructure. This can help Amtrak optimize the maintenance and repair of its assets, reduce downtime, and improve the overall performance of its operations.

Customer experience: Data analysis can be used to analyze data on customer satisfaction, such as survey results and feedback, to identify areas for improvement in the customer experience. This can help Amtrak tailor its services to better meet the needs and preferences of its customers.

REQUIREMENTS FOR SUBMISSION
> GitHub
> <XXXX>
> Write-up
> <XXXX>
> Syntax
> <XXXX>
> Data
> <XXXX>

FORMATTING
See "Assignment 1 - Descriptives" for a detailed list of assignment formatting guidelines. Also, assignment formatting guidelines can be found in the course document cache.

DATASET DETAILS (all data sets can be found here or here)
amtrak.csv

DATASET FIELDS

Code
StationName
Staffed
StationType
City
State
Zip
Country
Division
PublicPhone
TicketOffice
QuikTrak
MetrolinkTVM
CheckedBaggage
AmtrakExpress
HandleUSMail Lounge
UnaccompaniedChildren
AdvPayPPO
AmtrakTrains
Non-AmtrakTrains
DedicatedThruway
CoordinatedThruway
OtherTransportation
GroupCity
IATACode
IATAName
NonAmtrakTicketSale
TimeZone
ObserveDayLightSaving
CheckInWService
CheckInWOService
Elevation
Population