Project: Analysis of Airline Passengers and Identification of Factors Influencing Business Performance

Objective: Use dummy data to analyze the number of passengers in different airlines and identify factors that contribute to a successful airline business. Provide recommendations for improvement to other airlines based on the analysis.

Data Collection: Create or generate dummy data representing various airlines and their corresponding number of passengers. Include additional relevant variables such as flight routes, fares, customer ratings, on-time performance, and any other factors you deem important.

Data Cleaning and Preparation: Clean the data by handling missing values, removing duplicates, and ensuring consistency. Perform any necessary data transformations or standardizations.

Exploratory Data Analysis (EDA): Conduct EDA to gain insights into the data. Explore the distribution of passenger numbers, fares, and other variables across different airlines. Identify any patterns or relationships between variables that may influence passenger numbers.

Descriptive Analysis: Calculate descriptive statistics such as mean, median, and standard deviation of passenger numbers for each airline. Identify airlines with consistently high or low passenger numbers.

Comparative Analysis: Compare airlines based on factors such as customer ratings, on-time performance, and fares. Use visualizations, such as bar charts or box plots, to highlight differences among airlines.

Identify Factors Influencing Business Performance: Perform a regression analysis or use appropriate statistical techniques to determine which factors have the most significant impact on passenger numbers. This could include factors such as fares, customer ratings, flight routes, or on-time performance.

Recommendations for Improvement: Based on the analysis, provide recommendations to other airlines on how they can improve their business. For example, if low fares are found to be a significant driver of passenger numbers, suggest that airlines review their pricing strategies to remain competitive. If customer ratings are influential, recommend focusing on enhancing the customer experience.

Visualize Insights: Create visualizations to present your findings and recommendations. Use charts, graphs, and tables to effectively communicate the analysis results.

```python
import random
import pandas as pd

# Generate dummy data for airlines
airline_names = ['Buddha Air', 'Yeti Air', 'Saurya Air']
num_passengers = 100

# Generate dummy data for number of passengers
passenger_data = []
for airline in airline_names:
    passengers = [random.randint(50, 200) for _ in range(num_passengers)]
    passenger_data.extend(passengers)

# Generate dummy data for additional factors
customer_ratings = [random.uniform(3.0, 5.0) for _ in range(num_passengers * len(airline_names))]
on_time_performance = [random.uniform(70, 100) for _ in range(num_passengers * len(airline_names))]
fares = [random.uniform(4500, 9000) for _ in range(num_passengers * len(airline_names))]

# Create a pandas DataFrame to organize the data
data = pd.DataFrame({
    'Airline': [airline for _ in range(num_passengers) for airline in airline_names],
    'Passengers': passenger_data,
    'Customer Ratings': customer_ratings,
    'On-Time Performance': on_time_performance,
    'Fares': fares
})

# Print the generated dummy data
print(data.head())
```

```
      Airline  Passengers  Customer Ratings  On-Time Performance        Fares
0  Buddha Air          99          3.036215            71.875211  8816.827524
1    Yeti Air          60          4.404141            73.677768  8523.984558
2  Saurya Air          66          3.576332            97.363526  5684.502190
3  Buddha Air          58          4.979086            99.007954  7650.526041
4    Yeti Air         119          4.027123            92.261482  6983.489602
```

2)

```
# Handle missing values
data.dropna(inplace=True)  # Remove rows with any missing values

# Remove duplicates
data.drop_duplicates(inplace=True)

# Perform data transformations or standardizations
# Example: Standardize the 'Passengers' column using z-score normalization
mean_passengers = data['Passengers'].mean()
std_passengers = data['Passengers'].std()
data['Passengers'] = (data['Passengers'] - mean_passengers) / std_passengers

# Ensure consistency
# Example: Convert 'On-Time Performance' to a percentage
data['On-Time Performance'] = data['On-Time Performance'] * 100

# Print the cleaned data
print(data.head())
```

	Airline	Passengers	Customer Ratings	On-Time Performance	Fares
0	Buddha Air	-0.538603	3.036215	7187.521109	8816.827524
1	Yeti Air	-1.406481	4.404141	7367.776783	8523.984558
2	Saurya Air	-1.272961	3.576332	9736.352613	5684.502190
3	Buddha Air	-1.450987	4.979086	9900.795398	7650.526041
4	Yeti Air	-0.093538	4.027123	9226.148171	6983.489602

3)

```python
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Assume the cleaned data is stored in the 'data' DataFrame

# Summary statistics
print(data.describe())

# Distribution of passenger numbers
plt.figure(figsize=(10, 6))
sns.histplot(data=data, x='Passengers', hue='Airline', bins=10)
plt.title('Distribution of Passenger Numbers')
plt.xlabel('Passengers')
plt.ylabel('Frequency')
plt.show()

# Boxplot of fares
plt.figure(figsize=(10, 6))
sns.boxplot(data=data, x='Airline', y='Fares')
plt.title('Boxplot of Fares')
plt.xlabel('Airline')
plt.ylabel('Fares')
plt.show()

# Correlation heatmap
plt.figure(figsize=(10, 8))
corr = data[['Passengers', 'Customer Ratings', 'On-Time Performance', 'Fares']].corr()
sns.heatmap(corr, annot=True, cmap='coolwarm')
plt.title('Correlation Heatmap')
plt.show()
```

4)

```python
import pandas as pd

# Assume the cleaned data is stored in the 'data' DataFrame

# Calculate descriptive statistics of passenger numbers for each airline
statistics = data.groupby('Airline')['Passengers'].agg(['mean', 'median', 'std'])

# Identify airlines with consistently high or low passenger numbers
high_passenger_airlines = statistics[statistics['mean'] > statistics['mean'].mean()]
low_passenger_airlines = statistics[statistics['mean'] < statistics['mean'].mean()]

# Print the descriptive statistics and identified airlines
print("Descriptive Statistics of Passenger Numbers:")
print(statistics)
print("\nAirlines with consistently high passenger numbers:")
print(high_passenger_airlines)
print("\nAirlines with consistently low passenger numbers:")
print(low_passenger_airlines)
```

```
Descriptive Statistics of Passenger Numbers:
                mean    median       std
Airline
Buddha Air  0.032193  0.062235  1.033834
Saurya Air  0.020399 -0.037905  0.950533
Yeti Air   -0.052592 -0.049031  1.021575

Airlines with consistently high passenger numbers:
                mean    median       std
Airline
Buddha Air  0.032193  0.062235  1.033834
Saurya Air  0.020399 -0.037905  0.950533

Airlines with consistently low passenger numbers:
              mean    median       std
Airline
Yeti Air -0.052592 -0.049031  1.021575
```

5)

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

# Assume the cleaned data is stored in the 'data' DataFrame

# Bar chart of customer ratings
plt.figure(figsize=(10, 6))
sns.barplot(data=data, x='Airline', y='Customer Ratings')
plt.title('Customer Ratings by Airline')
plt.xlabel('Airline')
plt.ylabel('Customer Ratings')
plt.show()

# Boxplot of on-time performance
plt.figure(figsize=(10, 6))
sns.boxplot(data=data, x='Airline', y='On-Time Performance')
plt.title('On-Time Performance by Airline')
plt.xlabel('Airline')
plt.ylabel('On-Time Performance')
plt.show()

# Boxplot of fares
plt.figure(figsize=(10, 6))
sns.boxplot(data=data, x='Airline', y='Fares')
plt.title('Fares by Airline')
plt.xlabel('Airline')
plt.ylabel('Fares')
plt.show()
```

6)

```python
import pandas as pd
import statsmodels.api as sm

# Assume the cleaned data is stored in the 'data' DataFrame

# Select the relevant variables for the regression analysis
variables = ['Fares', 'Customer Ratings', 'On-Time Performance']

# Add a constant term for the intercept in the regression model
data['intercept'] = 1

# Fit the multiple linear regression model
model = sm.OLS(data['Passengers'], data[variables + ['intercept']])
results = model.fit()

# Print the regression results
print(results.summary())
```

Based on the analysis and the identified factors that significantly impact passenger numbers, here are some recommendations for other airlines to improve their business:

Pricing Strategy: If low fares are found to be a significant driver of passenger numbers, recommend that airlines review their pricing strategies to remain competitive. This may involve offering competitive prices, implementing dynamic pricing models, or introducing promotional offers to attract more passengers.

Customer Experience: If customer ratings are influential, suggest that airlines focus on enhancing the overall customer experience. This can include improving in-flight amenities, providing excellent customer service, offering personalized services, and addressing customer feedback promptly and effectively. By prioritizing customer satisfaction, airlines can attract more passengers and foster customer loyalty.

On-Time Performance: If on-time performance has a significant impact on passenger numbers, advise airlines to prioritize punctuality and operational efficiency. This can involve improving scheduling, optimizing flight operations, and minimizing delays. Keeping passengers informed about any schedule changes or disruptions can also help maintain their trust and satisfaction.

Marketing and Branding: Airlines can enhance their marketing and branding efforts to differentiate themselves in the market and attract more passengers. This can involve showcasing unique features or benefits, promoting special services or partnerships, and

leveraging digital marketing strategies to reach a wider audience. Building a strong brand image and conveying value propositions effectively can help airlines stand out in a competitive landscape.

Route Analysis: Analyze the flight routes and identify potential opportunities for expansion or optimization. This can include evaluating the demand for specific routes, identifying underserved markets, or exploring partnerships with other airlines to offer better connectivity. By strategically analyzing and adjusting their route network, airlines can tap into new markets and attract more passengers.

Data-Driven Decision Making: Encourage airlines to adopt a data-driven approach to decision making. By leveraging data analytics, airlines can gain insights into passenger preferences, market trends, and operational efficiency. This can help them make informed decisions regarding pricing, service enhancements, route planning, and overall business strategies.

It's important for each airline to assess its unique strengths, weaknesses, and market dynamics when implementing these recommendations. Tailoring strategies to their specific circumstances will lead to more effective improvements and better business outcomes.

```python
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Assume the necessary data and analysis results are available

# Bar Chart - Factors Influencing Passenger Numbers
plt.figure(figsize=(10, 6))
sns.barplot(x=['Fares', 'Customer Ratings', 'On-Time Performance'], y=[0.2, 0.5, 0.3])
plt.title('Factors Influencing Passenger Numbers')
plt.xlabel('Factors')
plt.ylabel('Impact')
plt.show()

# Pie Chart - Percentage of Passenger Numbers by Airline
plt.figure(figsize=(8, 8))
data['Airline'].value_counts().plot(kind='pie', autopct='%1.1f%%')
plt.title('Percentage of Passenger Numbers by Airline')
plt.ylabel('')
plt.show()

# Line Chart - Trend of Passenger Numbers over Time
plt.figure(figsize=(10, 6))
plt.plot(data['Date'], data['Passengers'])
plt.title('Trend of Passenger Numbers over Time')
plt.xlabel('Date')
plt.ylabel('Passenger Numbers')
plt.xticks(rotation=45)
plt.show()

# Scatter Plot - Relationship between Fares and Passenger Numbers
plt.figure(figsize=(10, 6))
sns.scatterplot(data=data, x='Fares', y='Passengers')
plt.title('Relationship between Fares and Passenger Numbers')
plt.xlabel('Fares')
plt.ylabel('Passenger Numbers')
plt.show()

# Table - Summary of Recommendations
recommendations = pd.DataFrame({
    'Recommendation': ['Review Pricing Strategy', 'Enhance Customer Experience',
                       'Prioritize On-Time Performance', 'Focus on Marketing and Branding',
                       'Analyze and Optimize Flight Routes', 'Embrace Data-Driven Decision Making'],
    'Factors to Focus On': ['Fares', 'Customer Ratings', 'On-Time Performance', 'Marketing and Branding',
                            'Flight Routes', 'Data Analytics']
})

plt.figure(figsize=(10, 4))
ax = plt.gca()
ax.get_xaxis().set_visible(False)
ax.get_yaxis().set_visible(False)
table = plt.table(cellText=recommendations.values, colLabels=recommendations.columns, cellLoc='center', loc='center')
table.set_fontsize(12)
table.scale(1.2, 1.2)
plt.title('Summary of Recommendations')
plt.axis('off')
plt.show()
```