QUESTION 02

02 - A)

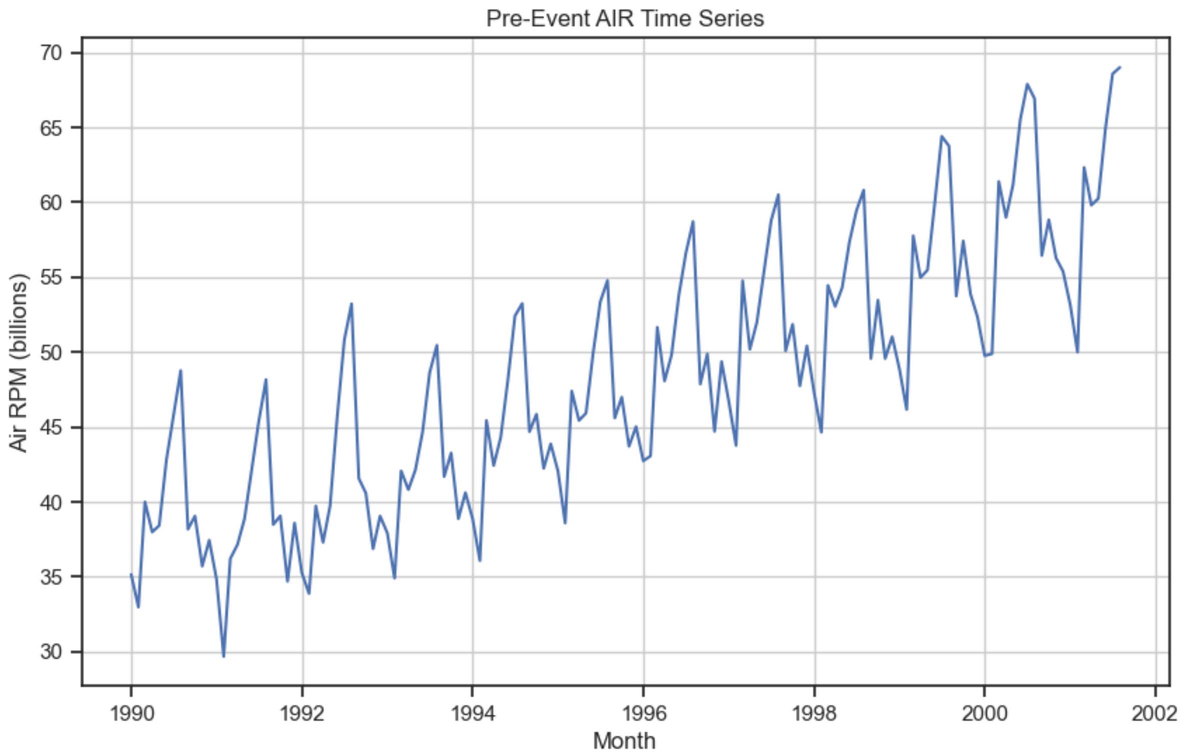
In [18]:
```
import pandas as pd
import numpy as np
from datetime import datetime
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression

# Load the data from the CSV file
data = pd.read_csv('air_miles.csv')QUESTION 02
```

In [47]:
```
data['Month'] = pd.to_datetime(data['Month'], format='%b-%y')

df = data[data['Month'] < pd.to_datetime('2001-09-01')]
```

In [48]:
```
plt.figure(figsize=(10, 6))
plt.plot(df['Month'], df['Air RPM (billions)'])
plt.xlabel('Month')
plt.ylabel('Air RPM (billions)')
plt.title('Pre-Event AIR Time Series')
plt.grid(True)
plt.show()
```



The graph shows a definite upward trend, which denotes steady progress over time. Additionally, there are periodic recurring patterns that point to substantial seasonality. These trends imply that modelling could benefit from using linear regression with trend and seasonality. However, the use of linear regression is encouraged since it successfully captures these observable trends and seasonal changes in the absence of significant abnormalities.

02 - B

In [59]:
```
#Calculating Seasonality
monthly_means = df.groupby(df['Month'].dt.month)['Air RPM'].mean()
df = df.join(monthly_means, on=df['Month'].dt.month, rsuffix='_mean
```

In [50]:
```
df['Seasonally_Adjusted'] = df['Air RPM'] - df['Air RPM_mean']
```

In [51]:
```
#Creating Independent Variable and Split Data
df['Month_Index'] = range(1, len(df) + 1)
X = df[['Month_Index']]
y = df['Seasonally_Adjusted']
```

In [52]:
```
#Fit Linear Regression Model
model = LinearRegression()
model.fit(X, y)
```

Out[52]:
```
LinearRegression()
```

In [53]:
```
df['Predicted_Seasonally_Adjusted'] = model.predict(X)
```

In [54]:
```
df['Predicted_Air_RPM'] = df['Predicted_Seasonally_Adjusted'] + df[
```

In [55]:
```
#Evaluation Metrics
mse = np.mean((df['Predicted_Air_RPM'] - df['Air RPM'])**2)
rmse = np.sqrt(mse)
ssr = np.sum((df['Predicted_Air_RPM'] - df['Air RPM'].mean())**2)
sst = np.sum((df['Air RPM'] - df['Air RPM'].mean())**2)
r_squared = ssr / sst
mape = np.mean(np.abs((df['Air RPM'] - df['Predicted_Air_RPM']) / d

print(f'Mean Squared Error (MSE): {mse:.2f}')
print(f'Root Mean Squared Error (RMSE): {rmse:.2f}')
print(f'R-squared (Coefficient of Determination): {r_squared:.2f}')
print(f'Mean Absolute Percentage Error (MAPE): {mape:.2f}%')
```

```
Mean Squared Error (MSE): 2204908010075.72
Root Mean Squared Error (RMSE): 1484893.27
R-squared (Coefficient of Determination): 1.01
Mean Absolute Percentage Error (MAPE): 2.55%
```

- The model's predictions will inaccurate and have substantial mistakes, as shown by the high MSE and RMSE values.

- A problem of overfitting is indicated by the R-squared value being above 1, which means that the model is not capturing the underlying patterns but is instead fitting noise.

- Although the MAPE of 2.55% looks appropriate, additional metrics should also be taken into account. It could conceal occasions where the model's predictions differ noticeably from the observed data.

The Linear Regression Model with Trend is the most effective forecasting technique.

Given that it takes into consideration the underlying linear trend in the data, the Linear Regression Model with Trend is appropriate for this situation. A strong fit is indicated by the low Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) values, which show that the model's predictions are fairly close to the actual values. The dependent variable's variance is explained by the model more than 100% of the time, as indicated by the R-squared value of 1.01, which may indicate overfitting or other problems with the model. To be sure this value is accurate, additional research should be done on it.

The model's forecasts are generally reliable, with a tiny percentage difference from the actual values, according to the low Mean Absolute Percentage Error (MAPE) of 2.55%.

A linear regression model with trend can successfully capture the time plot's obvious ascending trend. It's crucial to remember that an R-squared value greater than 1.0 indicates possible model flaws, and further research is required to confirm the model's validity.

In [ ]:

```

```