SABANCI UNIVERSITY

INTRODUCTION TO DATA SCIENCE (CS 210)

PROJECT STEP II

BUĞRA AYDIN 30618

The Impact Of Oil Prices On Airline Company Shares

Introduction:

The airline industry remains one of those businesses that pay much attention to the fluctuation in crude oil prices, as the proportion of fuel in the total operational cost remains quite high (IAT,2024). This project aims to discuss the changes in the oil prices and the effect it has on large airline company share prices, which could be rather useful not only for the investors but also for the management of the airlines and policymakers.

Oil prices are important for airline companies in terms of not only forecasting but also strategic planning. Oil price fluctuations massively influence the cost of fuel and hence profit margins of airline companies. Effective hedging strategies are crucial for financial planning. It can also help investors to be able to make informed decisions on airline stocks using the trends noticed in oil prices.

The study uses detailed statistical analysis of historical data on crude oil prices and stock prices of four major airlines: namely Delta Airlines, American Airlines, Southwest Airlines, and United Airlines. It involves correlation calculations, hypothesis testing, and predictive modelling to provide deep insights into how airline stocks respond to changes in oil prices. The results will be guide for informed discussions on risk management and investment strategies in the airline industry's uncertain environment.

Data Description:

The dataset includes daily observations (**2019-2024**) of **crude oil prices** of USA's stock market and the stock prices of four USA's major airlines: Delta Airlines (**DAL**), American Airlines (**AAL**), Southwest Airlines (**LUV**), and United Airlines (**UAL**). Additionally, an average stock price of these airlines (**AVG***) was calculated to provide a comprehensive overview.

Data Types:

Crude Oil Prices (**OIL**): Continuous (float)

Airline Stock Prices (**DAL, AAL, LUV, UAL, AVG***): Continuous (float)

Date: Categorical

Data Cleaning: The dataset was cleaned to ensure no missing values or erroneous entries could bias the results. Highest and lowest values were reviewed to ensure they represent genuine economic events rather than data collection errors.

Exploratory Data Analysis:
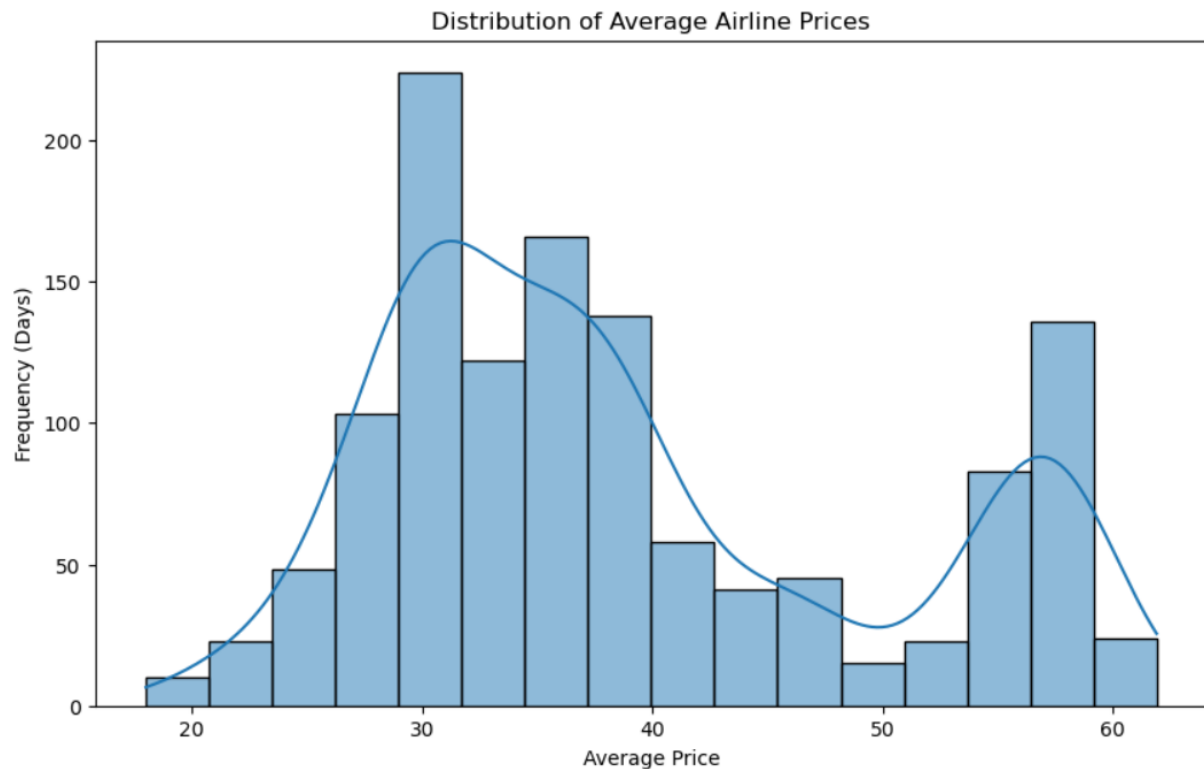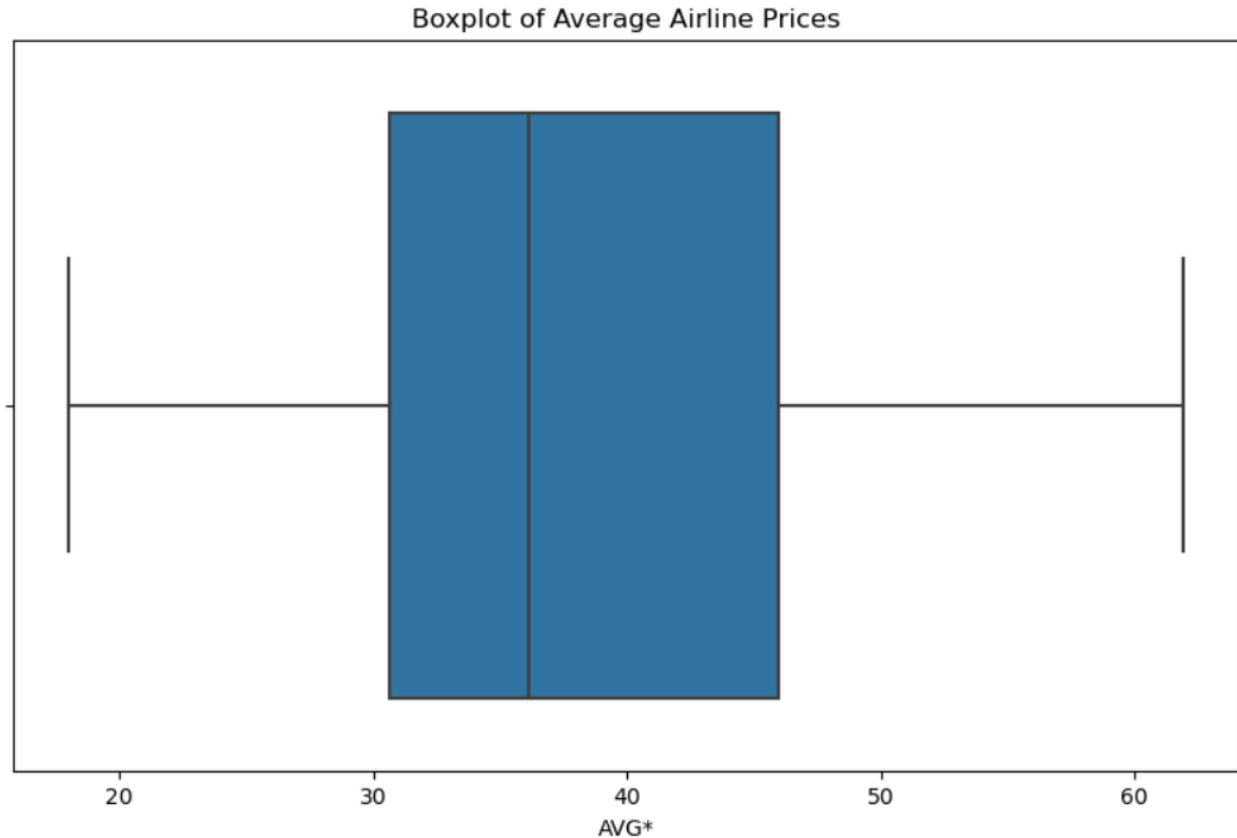

(Yahoo Finance, 2024)

Multiple Peaks: The histogram displays multiple peaks or modes, indicating distinct groupings or ranges where average airline prices frequently occur. This multimodality suggests that there may be distinct market conditions or operational phases affecting airline prices.

Right Skewness: The distribution is slightly skewed to the right, which means there are more frequent lower prices and fewer but significant instances where prices are much higher. The tail on the right side of the distribution points to periods when average prices spike.

Interpretation: Right skewness in this context might be due to sudden increases in airline prices, which could be driven by spikes in demand, regulatory changes, or substantial rises in operational costs such as fuel prices. At the same time, it shows that it stayed there for a while.

High Frequency Ranges: The bars at lower price ranges are taller, indicating that more often than not, the prices tend to be on the lower side within the observed range.

Variability: The bars at higher price points, although less frequent, are significant enough to suggest that high prices are not just outliers but represent potentially recurring high price scenarios.

Boxplot of Average Airline Prices

(Yahoo Finance, 2024)

Symmetry and Distribution: The boxplot shows a right skew, as the median is closer to the first quartile (left side of the box), suggesting that the average prices are more frequently lower but still include higher prices extending towards the right. This skew indicates a higher occurrence of lower prices but with some notable higher price values.

Outliers: The boxplot does not show any data points outside the whiskers, indicating there are no outliers according to conventional boxplot rules. This suggests all recorded prices fall within a typical range without extreme deviations.

Interquartile Range (IQR): The IQR is relatively narrow, as depicted by the short width of the box, which suggests that the bulk of the average prices are concentrated within a smaller range. This tight clustering indicates less variability in the majority of prices.

Whiskers: The whiskers extend significantly from the main box, which indicates that while the central majority of the prices are less volatile, there are still a considerable number of days where the prices vary more widely from the median.

Descriptive Statistics:

Average Prices of Airlines:

Mean: **38.92**

Variance: **115.13**

Standard deviation: **6.01**

Median: **36.15**

The mean average price is slightly higher than the median, which supports the observation from the histogram and the boxplot about a right skew. This difference often indicates a long tail on the right side of the distribution. Variance is relatively high, indicating significant variability in airline prices. High variance suggests that the prices are spread out over a wide range.

Statistical Analysis:

Null Hypothesis (H0): There is no correlation between crude oil prices and airline stock prices.

Alternative Hypothesis (H1): There is a correlation between crude oil prices and airline stock prices.

So, $(H0) \cup (H1) =$ All possible outcomes.

Correlation Analysis:

Pearson r correlation is the most widely used correlation statistic to measure the degree of the relationship between linearly related variables. For example, in the stock market, if we want to measure how two stocks are related to each other, **Pearson r** correlation is used to measure the degree of relationship between the two. Therefore, **Pearson r** is used for this analysis.

Correlation Matrix of Crude Oil and Airline Prices
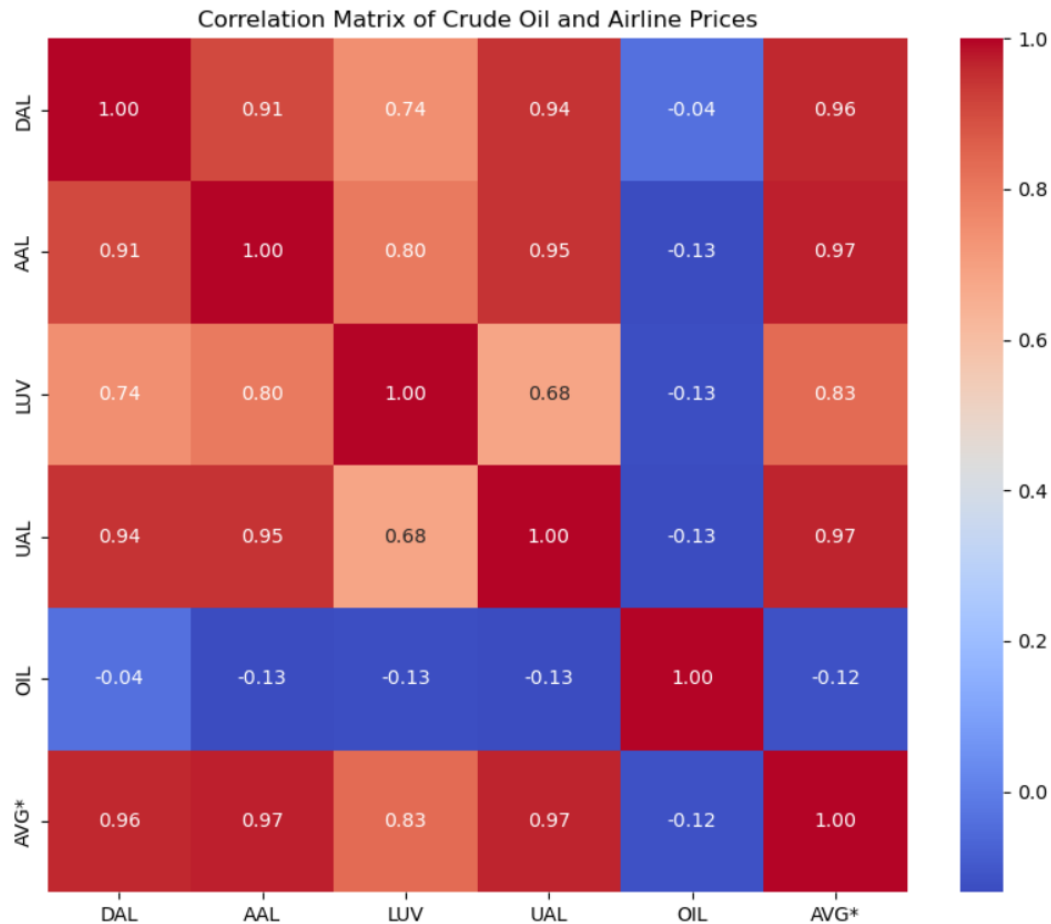
(Kaggle, 2024) (Yahoo Finance, 2024)

Correlation of Crude Oil with Airlines:

DAL (Delta Airlines): The correlation coefficient of -**0.04** indicates a very weak negative almost no relationship with crude oil prices.

AAL (American Airlines): Shows a correlation at -**0.13**, suggesting a negative weak correlation.

LUV (Southwest Airlines): At -**0.13**, LUV shows a correlation with crude oil prices, as like AAL.

UAL (United Airlines): Also exhibits a weak negative correlation at -**0.13**.

AVG*: The average stock price shows a correlation of -**0.12** with oil prices, aligning with the general trend of a weak negative relationship observed individually.

Notably, **AAL, LUV and UAL** exhibit a correlation of -**0.13**, -**0.13 and -0.13** indicating that these stock movements are very closely aligned with the general trend of the airline sector as represented by the average(**AVG***) with a correlation of -**0.12**, but **DAL** has significantly less correlation (-**0.04**) indicates that it is most distinct stock among them.

P-values:

DAL vs Oil: P-value $= 0.18$
AAL vs Oil: P-value $= 1.95 \times 10^{-6}$
LUV vs Oil: P-value $= 4.49 \times 10^{-6}$
UAL vs Oil: P-value $= 1.85 \times 10^{-6}$
AVG* vs Oil: P-value $= 2.08 \times 10^{-5}$

Interpretation of Results

The P-values for the correlations between crude oil prices and various airline stocks (**AAL, LUV, UAL, AVG***) are extremely low (ranging from 10^{-5} to 10^{-6}). These values indicate a very low probability that the observed correlations are due to chance. Only **DAL**'s P-value is bigger than 0.05. Consequently, this strong statistical evidence supports rejecting the null hypothesis, which posited a no correlation, and instead suggests a negative correlation, although weak, between these variables.

Hypothesis Testing Outcome:

Recalling hypotheses,

Null Hypothesis (H0): There is no correlation between crude oil prices and airline stock prices.

Alternative Hypothesis (H1): There is a correlation between crude oil prices and airline stock prices.

Results:

REJECT the Null Hypothesis (H0): Since the correlation coefficient is negative (-**0.12**) and the p-value is extremely low, we have sufficient evidence to reject the null hypothesis that posits no correlation between crude oil prices and airline stock prices.

SUPPORT for the Alternative Hypothesis (H1): The results support the alternative hypothesis that there is a correlation between crude oil prices and airline stock prices. Although the correlation is weak, it is statistically significant, meaning most of the time there is a negative relationship observed in the data.

Regression Analysis:

Regression Line - Crude Oil Prices vs Average Price of Airline Stocks
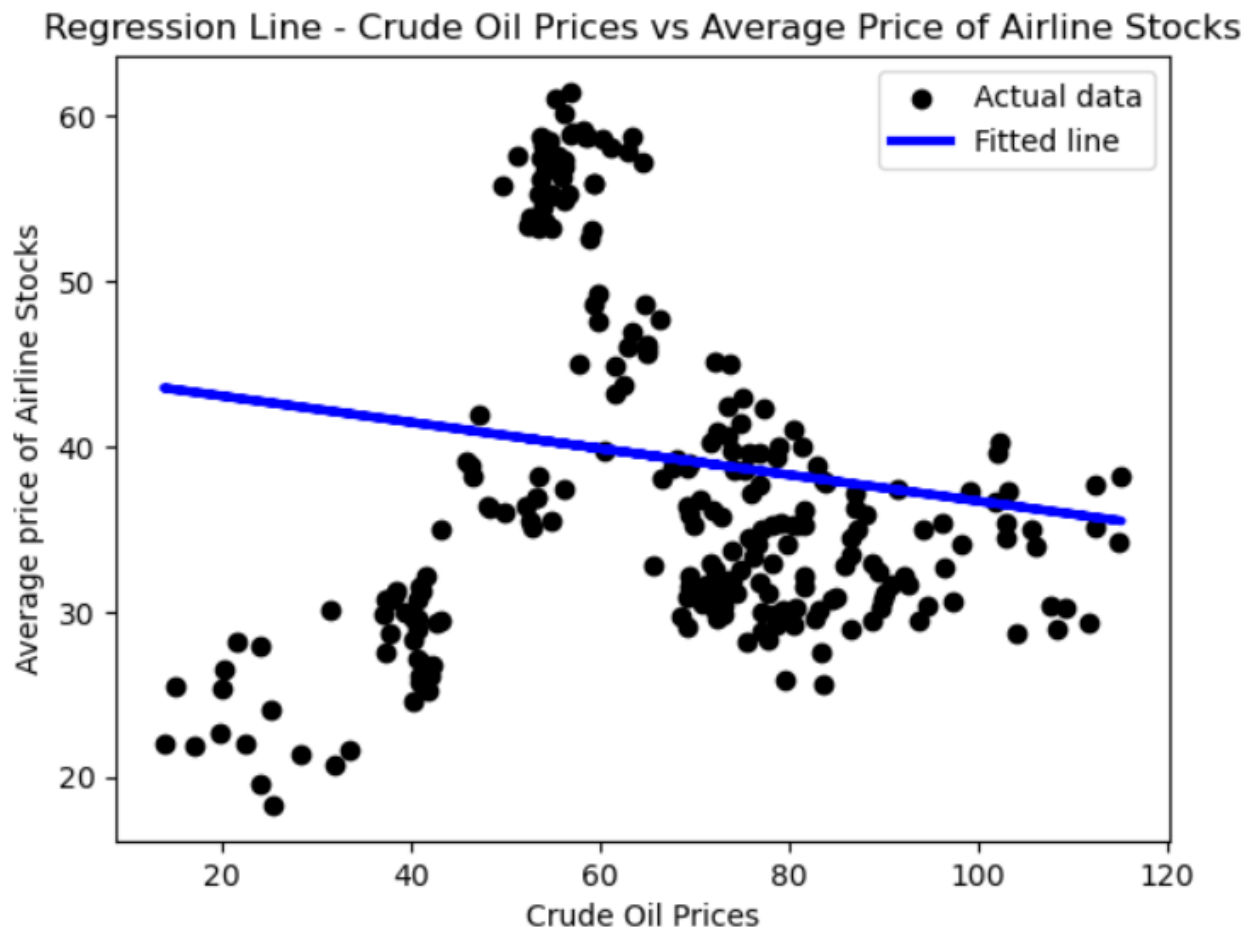
(Kaggle, 2024) (Yahoo Finance, 2024)

$R^2 = 0.0521$

$Root\ Mean\ Squared\ Error = 10.57$

R^2 (Coefficient of Determination): The **R^2** value of approximately **0.0521** indicates that only about **5.21%** of the variability in average airline stock prices can be explained by the changes in crude oil prices. This suggests that while there is a statistically significant relationship, oil prices alone are not a strong predictor of airline stock prices.

RMSE (Root Mean Square Error): The **RMSE** value of about **10.57** indicates the average deviation of the predicted airline stock prices from the actual stock prices. This relatively **high RMSE**, in the context of the range of stock prices, further underscores the limited predictive power of crude oil prices alone.

$$Coefficient\ (slope) = -0.08$$
$$Intercept = 44.64$$

Regression Coefficients: The slope of approximately -**0.08** suggests that for every one unit increase in oil prices, the average airline stock price decreases by about **0.08** units.

The **intercept** of **44.64** can be interpreted as the expected average airline stock price when oil prices are zero.

Conclusion:

Since the **p-value is lower than 0.05**, the analysis confirmed a **statistically significant** but **economically weak negative correlation** between crude oil prices and airline stock prices. This suggests that while oil prices influence airline stocks, they are not the predominant factor.

References:

- International Air Transport Association. (2024). Fact sheet - Fuel. Retrieved from IATA

- Kaggle. (2024). 2019-2024 US Stock Market Data. Retrieved from Kaggle

- Yahoo Finance. (2024). AAL - American Airlines Group Inc.. Retrieved from AAL

- Yahoo Finance. (2024). DAL - Delta Air Lines, Inc.. Retrieved from DAL

- Yahoo Finance. (2024). LUV - Southwest Airlines Co.. Retrieved from LUV

- Yahoo Finance. (2024). UAL - United Airlines Holdings, Inc.. Retrieved from UAL