Introduction

Colder days, cheaper tickets?

I like to travel back to my home country on beautiful days in summer, but my friends tell me this action is expensive, and you can travel to your home country in colder months of the year cheaper. So, in this project, I want to see if their claim is accurate or not.

In this project, I want to research and see the weather conditions and the temperature of the city influence the price of flights in the scope of North America. What are these changes, and are they increasing the prices or decreasing flight ticket prices, and are these changes alike? For example, if I buy a ticket on freezing days in winter, I should pay less than buying the same ticket in summer.

Used Data

Data Source

The project has two data sources, flight information, prices and the weather conditions that day on the scope of the project.

- **2018 Airplane Flights**, The flight of 2018 in North America. Important columns are 'PricePerTicket', 'Miles' and 'Quarter'. The 'Quarter' shows the time of 2018 when the flight happened, and the value is 1,2,3,4. The 'Miles' is the distance between Origin and Distance in mile metric. 'PricePerTicket' is the price for a ticket which the user paid.
    - License: CC0: Public Domain.
    - No missing data and clean dataset.
    - Data type: CSV
    - Metadata link: link
    - Source: Kaggle

- **Average day weather for 2018**, the average day weather of 2018 in North America. Important columns, 'DailyAverageDewPointTemperature' and 'DATE'. The 'DailyAverageDewPointTemperature' is the value of temperature in Fahrenheit, 'DATE' is the date of recorded weather:
    - License: CC0: Public Domain.
    - No missing data and clean dataset
    - Data type: CSV
    - Metadata link: link
    - Source: Kaggle

Pipeline

The pipeline which is written in python 3 download datasets from Kaggle. The datasets are clean. It maps the 'DATE' to the 'Quarter' value. Moreover, calculate the average weather in that quarter and the average of Price of the tickets per a mile in the quarter. The output of the pipeline is a CSV file which each row contains 'AvgTemperature' which is the average temperature in float, 'PricePerMiles' which is the average of Price of the tickets per a mile in float and the quarter number. The CSV file selected to store the output of pipeline because it produces a few number rows as output. At the end, remove unnecessary files.

Analysis

The pipeline produces 'AvgTemperature' and 'PricePerMiles' columns for each quarter. The question is wanting to see the weather is changes are the same way of changing the price or contrast. This need wants to calculate the correlation between these two columns. One way is the Pearson correlation coefficient (PCC) which measures linear correlation between two sets of data.

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \cdot \sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

Conclusions

The result of PCC is -0.65 which shows these two columns do not have correlation. So, the answer to the posed question "**Colder days, cheaper tickets?**" is no, the price of tickets increase in the cold temperatures. A reason for this phenomenon could be Christmas and holidays.

Also, this result is shown in the chart of two data.
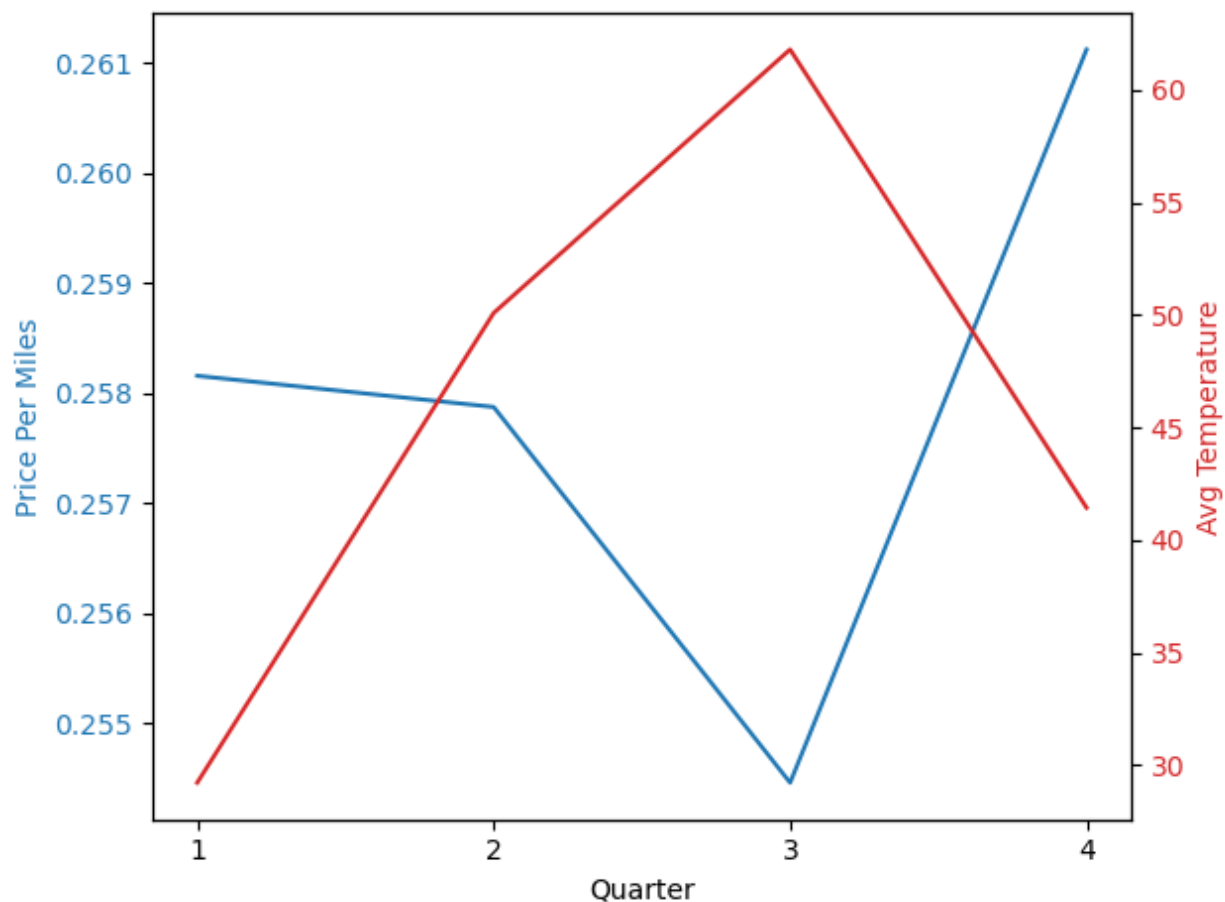
The result can be more accurate and shown in months of year, but the flight data source needs to have more accurate information about the flight time.