**Team 203 Feasibility Study:** Analysing Flight Ticket Price Factors

## Introduction

This research project will investigate the effects of flight duration, departure time and frequency on ticket price. Airlines will also be compared on average price of flights to certain destinations. Data will be sourced from either web scraping TripAdvisor or Kayak, or querying Skyscanner's API. The initial scope is limited to flights departing from Zurich Airport on April 11th, 2025.

## Motivation

Most of us can relate that, when trying to book a flight if we check the same flight the day after or from another company the prices fluctuate highly and sometimes, we miss the cheaper opportunities. This is a possible result of implementation of information technologies in pricing policies to maximize profits or minimize losses. Some research shows that there is an effect of departure time and flight duration on ticket prices (Szopiński et al., 2015 [1]; Mumbower et al., 2014 [2]). We would like to build upon this research and gain a comprehensive understanding of the factors affecting flight ticket prices.

## Background

In 2023 there were 67,300 flight routes worldwide (Aviation Benefits Without Borders, 2024 [3]). In 2023, 4.4 billion passengers travelled on 35.3 million flights. This included 1.8 billion passengers on international flights and 2.6 billion passengers on domestic flights.

## Research Questions

1. How do different airlines compare in terms of average ticket prices for flights to the same destination on the same day?
2. What is the impact of departure time on ticket prices for short-distance flights (flights under four hours) within a one-day observation period?
3. How does the frequency of available flights on a given route within one day affect the average ticket price?
4. What is the relationship between flight duration and ticket prices?

## Data

In order to answer the research questions, we will collect details of all the flights leaving Zurich airport on a randomly selected day, April 11th. Key features will include flight ID, departure time, date, flight duration, airline carrier name and ticket price. As we want to focus on short-distance flights, we will only consider flights lasting no more than four hours. After preliminary research, we decided that data will be sourced from either web scraping TripAdvisor[4][5] or Kayak[6][7], or querying Skyscanner's API[8][9].

## Expected Outcomes

1. **Comparative analysis of airline pricing strategies**: By comparing airlines on average ticket prices for identical routes on same day, we hope to uncover pricing trends and variations.

2. **Impact of departure time on pricing**: Our study will evaluate how different departure times influence ticket costs, potentially revealing peak-hour surcharges or discounts for less favourable flight times.
3. **Effect of flight frequency on ticket prices**: We expect to find correlations between flight frequency and pricing, which could indicate whether higher availability results in lower prices due to competitive pressures.
4. **Relationship between flight duration and ticket prices**: The study will assess whether shorter flights are priced proportionally to their duration or if pricing follows a different pattern based on airline strategies.
5. **Insights for consumers**: By identifying key trends in ticket pricing, we aim to provide travellers with practical recommendations for securing the best fares.

## Risks & Limitations

The major flight websites we explored for suitability were Google Flights, Skyscanner, TripAdvisor, Kayak and Kiwi. Respective robots.txt files showed that web scraping is disallowed for Google Flights[10], Skyscanner[11] and Kiwi[12]. Research also made clear that there is no free, public API available for Google Flights[13], TripAdvisor[14], Kayak[15] and Kiwi[16].

The challenges for web scraping flight websites are similar to the 5 V's of Big Data: There are massive amounts of data (volume) uploaded dynamically (velocity) and we need to develop an algorithm to extract only the data we want. In addition, there are many variables (variety) from which we need to extract only the one's that we are interested in. Another challenge we may face is extracting the data from one airport to "anywhere". We know that this option exists in Skyscanner, however, we are not sure if it is included in the API.

Due to time limitations, we decided to include data for only one day. This limitation of our study will prevent us from generalizing our results, since we won't investigate further important confounders such as seasonality, special holidays and weekend-weekday differences.

## Backup Plan

We have identified three potential data sources. If we face troubles in data wrangling from the first source, we will reattempt from the second or the third. If the option of flights departing to "anywhere" is not available, we will curate a list of all destination airports connected with Zurich Airport and scrape/ query for each destination in turn.

## References


[1] Tomasz Stanisław Szopiński, Robert Nowacki (2015) : The influence of purchase date and flight duration over the dispersion of airline ticket prices, Contemporary Economics, ISSN 2084-0845, Vizja Press & IT, Warsaw, Vol. 9, Iss. 3, pp. 353-366, https://doi.org/10.5709/ce.1897-9254.174
[2] Stacey Mumbower, Laurie A. Garrow, Matthew J. Higgins (2014): Estimating flight-level price elasticities using online airline data: A first step toward integrating pricing, demand, and revenue optimization, Transportation Research Part A: Policy and Practice, Volume 66, Pages 196-212, ISSN 0965-8564, https://doi.org/10.1016/j.tra.2014.05.003.
[3] Aviation Benefits Without Borders- Full Report, 2024, ATAG, https://aviationbenefits.org/media/e5ynn4x0/abbb2024_full_report.pdf
[4] https://www.tripadvisor.com/CheapFlightsHome, TripAdvisor
[5] https://www.tripadvisor.com/robots.txt, TripAdvisor
[6] https://www.kayak.com/flights, Kayak
[7] https://www.kayak.com/robots.txt, Kayak
[8] https://developers.skyscanner.net/docs/flights-live-prices/overview, Skyscanner
[9] https://developers.skyscanner.net/docs/flights-indicative-prices/overview, Skyscanner
[10] https://www.google.com/robots.txt, Google
[11] https://www.skyscanner.com/robots.txt, Skyscanner
[12] https://www.kiwi.com/robots.txt, Kiwi
[13] https://duffel.com/blog/google-flights-api, Krista Wymenga, Last Updated in 11/2021
[14] https://tripadvisor-content-api.readme.io/reference/terms-of-use, Last Updated on 12/12/2022
[15] https://affiliates.kayak.com/SignUp, Kayak
[16] https://kiwicom.github.io/margarita/docs/guide-tequila-api-key, Filip Messa, Last updated on 3/26/2019