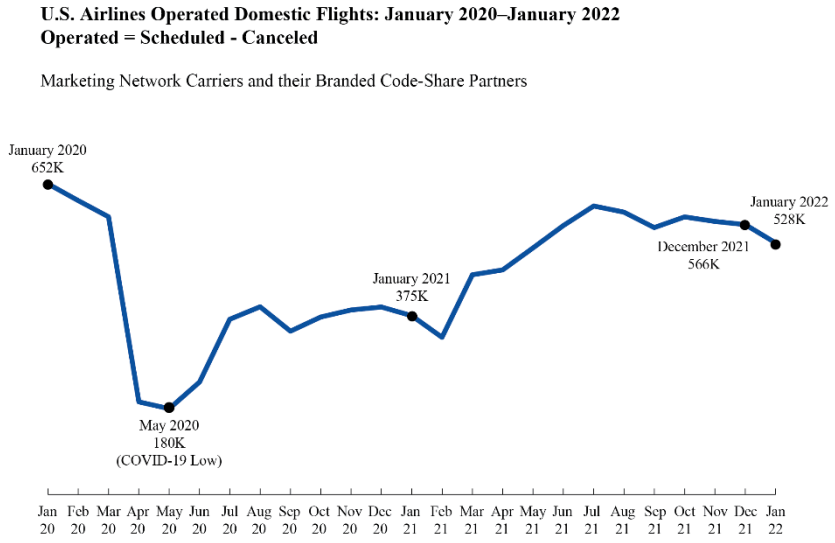
Michigan Ross Business+Tech Datathon 2023

Problem Statement: Airline Industry

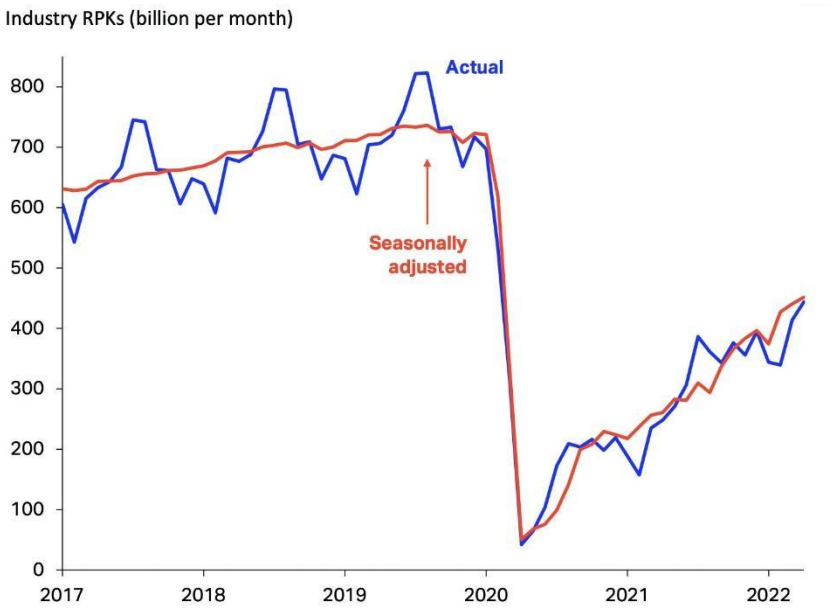
Background

The US Airline industry needs no introduction. US was home to the first commercial passenger flight and remains a pillar of the economy – accounting for 5% to US GDP, employing 10 million jobs and contributing to $1.3 trillion in economic activity.



(Source: https://www.transportation.gov/briefing-room/air-travel-consumer-report-january-2022-numbers)

However, from 2020, due to the huge impact of COVID-19, the airline industry reached its lowest industry point in recent years. Since then the industry has seen a rebound although several challenges remain.



(Source: https://ca.news.yahoo.com/cancelled-flights-disrupted-vacations-frayed-121154183.html)

Airline industry is in uncertain times.

For Michigan Ross Datathon 2023, you are provided with data on airline industry. You are asked to explore and analyze this data to answer relevant business questions.

Your Task

Your goal is to analyze the provided datasets (described in detail below), potentially in combination with supplementary public information, to ask relevant business questions and answer them. Please note that we are not providing specific question prompts by design. You are asked to explore what questions can be asked and answered using the provided data.

You are asked to pose your own question, explore your own solution approach, and provide answers/solutions using the available datasets in the available time. What is important is the insightfulness and depth of your conclusions and analysis. You need not be comprehensive; quality data analysis will be rewarded over the breadth of the question posed. You however must look at your task as broader than just data analysis – **you should look for a business question and consider the business context**. Note that thoughtfully prepared data visualizations are an integral part of your data exploration and analysis.

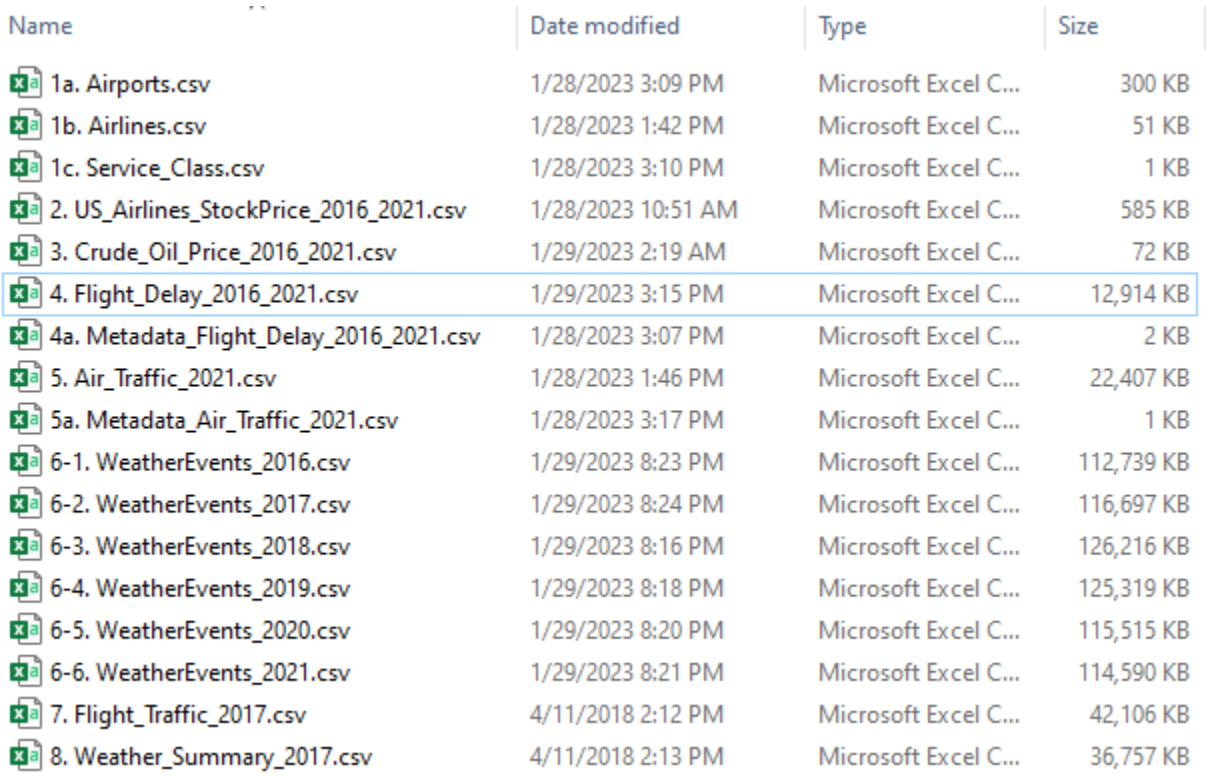
Creativity in formulating your own well-thought-out question or analysis approach generally has a positive effect on judges' assessment of your submission; however, it should not be at the expense of analytical depth, precision, and rigor, which are far more important.

Your analysis should be well communicated to a non-technical audience, potentially with the help of appropriate visualizations.

Datasets

Datasets are provided in CSV format. CSV files can be easily opened with all data analysis tools including MS Excel. Please note that the datasets are **only partially cleaned** and you may need to do some data cleaning before you can proceed to data analysis. There may be some redundant and duplicate data as well – so use your judgment when combining different data sets.

You are provided with the following datasets (see image below). Note that while most column headers are self explanatory, additional metadata files are provided (indicated by file number 4a,5a etc.) to explain the column headers.



Name	Date modified	Type	Size
1a. Airports.csv	1/28/2023 3:09 PM	Microsoft Excel C...	300 KB
1b. Airlines.csv	1/28/2023 1:42 PM	Microsoft Excel C...	51 KB
1c. Service_Class.csv	1/28/2023 3:10 PM	Microsoft Excel C...	1 KB
2. US_Airlines_StockPrice_2016_2021.csv	1/28/2023 10:51 AM	Microsoft Excel C...	585 KB
3. Crude_Oil_Price_2016_2021.csv	1/29/2023 2:19 AM	Microsoft Excel C...	72 KB
4. Flight_Delay_2016_2021.csv	1/29/2023 3:15 PM	Microsoft Excel C...	12,914 KB
4a. Metadata_Flight_Delay_2016_2021.csv	1/28/2023 3:07 PM	Microsoft Excel C...	2 KB
5. Air_Traffic_2021.csv	1/28/2023 1:46 PM	Microsoft Excel C...	22,407 KB
5a. Metadata_Air_Traffic_2021.csv	1/28/2023 3:17 PM	Microsoft Excel C...	1 KB
6-1. WeatherEvents_2016.csv	1/29/2023 8:23 PM	Microsoft Excel C...	112,739 KB
6-2. WeatherEvents_2017.csv	1/29/2023 8:24 PM	Microsoft Excel C...	116,697 KB
6-3. WeatherEvents_2018.csv	1/29/2023 8:16 PM	Microsoft Excel C...	126,216 KB
6-4. WeatherEvents_2019.csv	1/29/2023 8:18 PM	Microsoft Excel C...	125,319 KB
6-5. WeatherEvents_2020.csv	1/29/2023 8:20 PM	Microsoft Excel C...	115,515 KB
6-6. WeatherEvents_2021.csv	1/29/2023 8:21 PM	Microsoft Excel C...	114,590 KB
7. Flight_Traffic_2017.csv	4/11/2018 2:12 PM	Microsoft Excel C...	42,106 KB
8. Weather_Summary_2017.csv	4/11/2018 2:13 PM	Microsoft Excel C...	36,757 KB

Caution: you must first spend enough time to understand the dataset. Datathon Canvas discussion forums are a great place to ask and answer questions about the data.

Warning: despite our best efforts, the weather events data (file 6-1 through 6-6) remain too large to fit in an MS-Excel Worksheet – so plan accordingly. **Please note that you need not use every bit of data provided – use the part of the data that speaks to you, that sparks curiosity in you, that lends insights to you.**
Note: No further information will be provided other than the data files above and this problem statement. Teams must make their own assumptions and document them. Feel free to search online for additional information and augment the information provided. You can also post in Datathon Canvas discussion forum but we will limit our replies to clarifications about the dataset.

Additional Datasets

You are welcome to scour the Web for additional datasets/information to supplement your analysis. All additional data used should be public and their source must be documented in the report.

Submission: Content

You are expected to submit two separate deliverables: a report and a presentation file.

Report

Your primary submission should be in the form of a report that should have following main components:

1. **Executive Summary** – A one to two page summary of your key insights and proposed solution.
2. **Non-Technical Solution Summary** – What is the question that your team set out to answer? What were your key findings, and what is their significance? You must communicate your insights clearly – summary statistics and visualizations are encouraged if they help explain your thoughts.
3. **Technical Exposition** – What was your methodology/approach towards answering the questions? Describe your data manipulation and exploration process, as well as your analytical and modeling steps, if any.
4. **Code** – If you used a programming platform like R or Python then please include all relevant code that was used to generate your results. If you used a non-code based analysis platform (e.g. Excel) then your Technical Exposition section should have sufficient details to communicate your analysis steps.

Additional information (e.g. roadblocks encountered, caveats, future research areas, and unsuccessful analysis pathways) may be placed in an appendix.

Judges will be evaluating your work without your team there to explain it; therefore, your submission must "**speak for itself**". It need not be polished to the level of a final product, but do ensure that your main findings are clear and that any visualizations are functionally labeled and interpreted.

Note that since we don't provide a prompt, you are free to choose your audience. You may look at the data from the point of view of airlines, customers, regulatory bodies or any other stakeholders. However, you should clearly document your audience and the value you seek to provide to them.

Presentation

In addition to a report, you are asked to build and submit a presentation file. This is the file you will use to present your solution to the expert panel of judges should you be chosen as one of the 6 finalist teams. **Note that the deadline for presentation file is later than the deadline for the main report. You should first build and submit your report and then work on building the presentation file.**

Submission: Evaluation

You will be evaluated based on your Report, as follows:

- **Non-Technical Solution Summary: <u>This is the most important part of your report.</u>**
    - **Insightfulness of Conclusions**. What is the question that your team set out to answer, and how did you answer it? Are your conclusions precise and nuanced, as opposed to blanket (over)generalizations? Is your proposed solution well developed not just from a data/technical point of view but also business point of view? Are you going beyond the obvious? If proposing a business solution, are you considering risks, challenges, implementation plans and so on.
- **Technical Exposition:**
    - **Wrangling & Cleaning Process.** Did you conduct proper quality control and handle common error types? How did you transform the datasets to better use them together? Please describe your process in detail within your Report.
    - **Investigative Depth.** How did you conduct your exploratory data analysis process? What other hypotheses tests and ad-hoc studies did you perform, and how did you interpret the results of these? What patterns did you notice, and how did you use these to make subsequent decisions?
    - **Analytical & Modeling Rigor.** What assumptions and choices did you make, and what was your justification for them? How did you perform feature selection? If you built models, how did you analyze their performance, and what shortcomings do they exhibit? If you constructed visualizations and/or conducted statistical tests, what was the motivation behind the particular ones you built, and what do they tell you?

Submissions: Format

Reports can be produced using any tool you prefer; however, your report MUST be in a universally accessible and readable format (HTML, PDF, PPT, Word etc). It must not require obscure dedicated software to open.

Code, if any, should be submitted in a single zipped collection of files separate from your report.

You will be asked to make your submissions to a Canvas assignment. There are separate assignments for submitting your report and your presentation file. Note that deadlines are strictly enforced and any significant delay in submission will result in your submission being rejected.

Tips & Recommendations

We recommend that you follow a low-hanging-fruit first approach. While it is tempting to go for big and bold ideas, remember that you have limited time and even essential tasks like data cleaning and joining tables are likely to take significant time. We strongly recommend that you take care of low, effort high return tasks first before going for more in-depth analytics. In particular, effective data visualization are an efficient, high return, low(er) effort target to finish first. It is important that you do not get stuck in a complex technical workflow and lose track of not just the time but also your overall analysis objective. If something is taking too long, then it's okay to shelve it for the moment and move on to more easily executable tasks. Prioritize simple, easy, doable tasks first.

We recommend that your team not try to learn new tools if possible; instead, leverage your existing skills to extract as much insight from the data as you can. You will find that even basic tools like Excel and Tableau, if used well, can be quite effective in a time bound competition like this where speed is essential. Further, a good business narrative and interpretation from a simple analysis is likely more valuable than complex technical analysis that does not lend itself to (or does not leave time for) a compelling business insight.

Note that your report will be evaluated for its overall communication, structure and narrative – and not just on your technical work. In fact**, if your report does not adequately communicate the story coming out from your technical work then all your work is essentially wasted**. Your report should present a compelling story and that story should be clear even with a cursory look at your report.

We STRONGLY encourage you to start typing up your final submission AT LEAST two to three hours before the submission deadline. In the past, many teams have spent a lot of time conducting great analyses, only to realize that they left almost no time for actually writing up and presenting their results. This cannot be stressed enough – quality data analysis that is incomplete or poorly presented will NOT win the competition.

Ask for Help

Do not hesitate to ask for help. You will have access to a Canvas discussion board where you can ask questions and the Datathon organizers will respond as appropriate. Representatives from sponsor companies will be offering Office Hours where you can get their advice. Datathon organizing team will be available through the day in case you have any technical or logistical issue – just let us know and we will do our best to assist.

If you wish to ask a question privately (i.e. not in the public Canvas Discussion Thread) then please send an email to Sanjeev (sankum@umich.edu).

That's it. We hope that you will have a wonderful experience in the Michigan Ross Datathon 2023 and that it will prove a fun learning experience for you.

Appendix: Inspiration

Here are the winners of past Datathons. Look at these smiling faces and giant checks. This year you could be the ones smiling and holding a giant check!