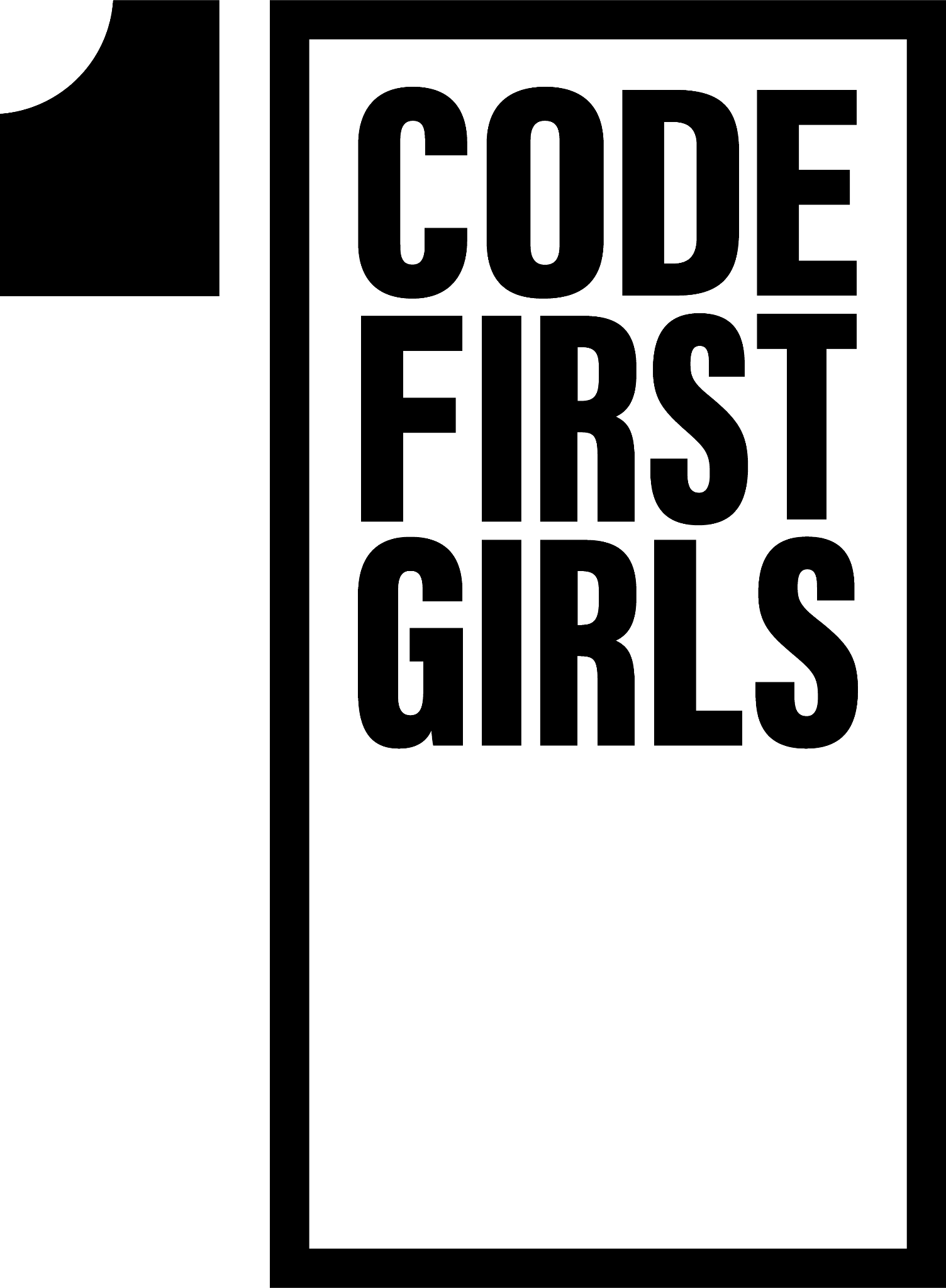
**CFG Nanodegree - Data 1 - Group Project**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

Homework Week Two

**Data Analysis Project Report**



**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**By Ashruti, Geneva, Tsamaija, Saweena, Zsoka, Yolanta**

**Contents**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

1. **Our Project Idea**
2. **Our Project Outline**
   1. **Details on Data Research and Analysis**
   2. **Chosen Industry and Areas**
   3. **Project Questions**
   4. **Data Sources**
   5. **Team Approach**

**1. Our Project Idea**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

When considering current events in contemporary society, it is evident that climate change and concerns for the welfare of the environment are at the forefront of socioeconomic issues. The aviation industry is one component of the economical landscape which contributes to environmental changes; the release of emissions, namely carbon dioxide, through air travel is an example of the way in which airlines can impact the environment. Air travel is a very time-efficient method of covering large distances, thus with the increasingly fast pace of modern life, for many wealthy countries, frequent flying has become a common practice. Every year, the United States of America (USA) has the highest number of airline passengers of any country in the world. In 2019, their passengers accounted for one fifth of all airline passengers worldwide, hence it is clear that finding even small ways to reduce the emissions of any airline in the USA will make a significant difference to emissions. As air travel is such a significant facet of the current transportation industry, this ‘Code First Girls’ Data Nanodegree project shall navigate through the Aerospace/Defence and Environment pathways by exploring the original idea of comparing the environmental impact of the USA’s main airlines. This includes the ten airlines which have almost a 90% share of the domestic flight market in the USA: American, Southwest, Delta, United, Spirit, Alaska, JetBlue, SkyWest, Frontier and Hawaiian Airlines. Furthermore, this project aims to investigate correlations between emission data and financial parameters such as ticket price or airline market share.

**2. Our Project Outline**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

The following information covers the important aspects of our project, detailing the data research and analysis we will undertake, our chosen industry areas, questions we aim to answer throughout our project, the data sources we will use and finally the approach we will take as a team.

1. **Details on Data Research and Analysis**

For this project, we aim to utilise a myriad of data research and analysis methods. One of the ways in which we intend to do this is through statistical analysis to determine causal - or correlational - relationships between our data. We will gather data from reputable sources, utilising API’s where possible, and process the data in order to analyse and interpret it. This analysis will be quantitative with the results being visualised using Python packages such as Matplotlib, to aid our interpretation of the data.

Moreover, this project offers the opportunity to carry out linear regression/time series analysis based on a financial parameter or an alternate variable such as fleet size. This regression analysis could be utilised to predict the emissions output of a flight based upon the ticket price, or even predict how emissions may increase over time. If there is a linear relationship between a financial metric - such as the price of a ticket or the market share of an airline - compared to the emissions produced by an airline, then this will be finalised as our project progresses.

1. **Chosen Industry and Areas**

There is no one single topic which encompasses the entirety of our project, instead it covers a range of areas which overlap and feed into each other:

* Aerospace: Civil Aviation
  + Considering the nature of our project idea, it is evident that our project heavily focuses on the aviation industry, specifically that offered by the USA and emissions produced by the chosen airlines.
* Environmental Issues
  + As aforementioned, there is some sort of relationship between air travel and its impact on the environment due to the production of emissions, thus the environment plays a factor in the output of our project.
* Business: Finance
  + There may be a correlation between plane ticket prices/airline market share and emission production, thus if our project continues in this direction it touches upon the financial aspects of the aviation industry.

1. **Project Questions**

This project aims to investigate the environmental impact of USA-based airlines while also reporting the factors which promote and predict low emissions and a reduced carbon footprint. To achieve this analysis, the following research questions have been proposed:

1. *How is a passenger’s carbon footprint impacted by the airline used?*

Using data from the Bureau of Transportation Statistics (BTS) on the monthly fuel consumption of the USA’s ten largest passenger airlines from 2010 - 2019, we will be able to calculate the total carbon emissions of each airline via fuel combustion each year. The BTS also provides data on the number of passengers transported by each airline each year in the Air Carrier : T-100 Segment, so combining the two datasets we can calculate the average carbon footprint per passenger.

Time and resource permitting, we would like to develop a carbon footprint calculator as an extension to this question, allowing individuals to dynamically calculate the emissions related to their planned journey.

1. *Investigating the link between total emissions of each airline via fuel combustion and the fleet profile/breakdown of each airline (2010 - 2019).*

We will perform an in-depth analysis of the fleet of each airline, combined with the previously calculated yearly fuel use to identify what impact the changing fleets of each airline over time had on an airlines fuel consumption.

1. *Investigating a possible correlation between average ticket prices and emissions per airline.*

The aim is to determine whether the carbon footprint of air travel is reflected in the prices passengers pay for their flights in the United States. For this, we are going to use our previously calculated emission data and the air fare information published by the U.S. Department of Transportation.

1. *Determining whether there is a correlation between company revenue, profit and emissions?*

Data analysis to investigate the relationship, if any, between airline revenue and profit and carbon emissions. We aim to use the BTS statistics records to gather financial data related to different airlines and the emission data gathered from each airline.

1. **Data Sources**

A wealth of historical aviation data can be found online, provided by The United States Department of Transportation. Their Bureau of Transportation Statistics (BTS) collects data from airlines via survey forms and stores the responses in online data tables. This data will form the basis for our data analysis project.

The following sources are not an exhaustive list, but are significant sources we have discovered to aid the development of our project:

* <https://www.transtats.bts.gov/> - the graph presented on this page has been used to pick our top 10 airlines to study. It is significant to our project as generating this statistic from scratch would be challenging and seems like an unnecessary step.
* <https://www.transtats.bts.gov/fuel.asp?pn=1> - this provides us with the fuel in gallons consumed each year by each airline e.g. Airline Fuel Cost and Consumption (American Airlines - Scheduled, January 1990 - June 2021).
* <https://www.transtats.bts.gov/DL_SelectFields.asp?gnoyr_VQ=GDM&QO_fu146_anzr=Nv4%20Pn44vr4> - provides a wide range of downloadable csv data tables with customisable columns including information on Carriers (Airlines), Aircraft, Time periods, Flight origins/destinations and other useful summary fields such as flight time or distance between airports.
* <https://www.carbonresponsible.com/carbon-responsible-api-overview/> - The Carbon Responsible API for carbon emissions measurement and enables the conversion of fuel, energy, travel, freight, water and paper use into carbon emissions from usage data held by companies in their finance or operational datasets. It uses DEFRA conversion factors and methodology to deliver the conversion of activity data into carbon emissions for reporting or voluntary management and reduction programmes.
* <https://www.airfleets.net/flottecie/American%20Airlines.htm> - provides the fleet details of American Airlines with information such as regarding the aircraft, the active or parked status of current fleet, details on stored or scrapped fleet, history and age.
* <https://data.transportation.gov/Aviation/Consumer-Airfare-Report-Table-1-Top-1-000-Contiguo/4f3n-jbg2> - consumer airfare report of top 1,000 contiguous state city-pair markets. Information on contiguous state city-pair markets that average at least 10 passengers per day [here](https://data.transportation.gov/Aviation/Consumer-Airfare-Report-Table-6-Contiguous-State-C/yj5y-b2ir) and domestic airline consumer airfare report [here](https://www.transportation.gov/policy/aviation-policy/domestic-airline-consumer-airfare-report).
* <https://www.transtats.bts.gov/AverageFare/> - average domestic airline itinerary fares; data that is searchable by year and quarter but does not list by airlines.
  + Original source of data:

<https://www.bts.gov/topics/airlines-and-airports/origin-and-destination-survey-data> - Origin and Destination Survey (DB1B) is a 10% sample of airline tickets from reporting carriers. Data includes origin, destination and other itinerary details of passengers transported.

* <https://www.bts.gov/topics/airlines-and-airports/quick-links-popular-air-carrier-statistics> - different types of American aviation related statistics.
* <https://www.easa.europa.eu/domains/environment/icao-aircraft-engine-emissions-databank> - information on emissions produced per engine.
* <https://www.eesi.org/papers/view/fact-sheet-the-growth-in-greenhouse-gas-emissions-from-commercial-aviation> - fact sheet on the growth in greenhouse gas emissions from commercial aviation in 2019. Includes information such as airline energy intensity and emissions, aviation emissions, regulating aircraft emissions, carbon emissions from international aviation, historically resilient growth and projections and global economic growth in the aviation industry.

1. **Team Approach**

Our team approach is one where we ensure we recognise each other's strengths while also being aware of each other's weaknesses and then to utilise this knowledge to determine each other's roles for the ongoing development and success of this project. This has been done by each team member performing a SWOT analysis, to identify our respective strengths, weaknesses, opportunities and threats. We have already implemented this information in our project so far by delegating tasks in accordance to each other’s level of comfort and expertise on a specific task, and hope to continue the workload this way as the project progresses.

Moving forward, we will be utilising agile development techniques, such as by constantly revising the progress of the project and implementing new tools and methods when required, while always recognising that there is room for improvement and adaptability as it is an iterative process. Furthermore, scrum methodology will be used for project management, such as by having at least two weekly meetings through voice chat on Discord alongside daily messaging on Discord to ensure that all team members are on track to achieve the goals assigned during the weekly meetings. These meetings will allow us to assess future tasks and therefore create a prioritised list of tasks, as well as allow us to work on resolving any other issues that have arisen since our last meeting that could not be easily discussed over Discord.

Alongside utilising the voice chat and messaging features on Discord, we have decided to use Slack as a formal reminder platform for meetings, as well as [Google Colab](https://colab.research.google.com/?utm_source=scs-index) for rough draft code and GitHub as we progress our work on the project. After using Google Colab for rough draft code, we then intend to utilise a Jupyter notebook file in GitHub to work on the final code that will be submitted at the end of the project.

**TEAM SWOT ANALYSIS**

|  | **STRENGTHS** | **WEAKNESSES** | **OPPORTUNITIES** | **THREATS** |
| --- | --- | --- | --- | --- |
| **Tsamaija** | Aerospace knowledge  Str, dict and list comprehension (this is only a relative strength)  Presenting | API connections | Looking forward to data analysis (incl graphs, charts etc) | Low time availability on weekdays. |
| **Geneva** | Data visualisation with Matplotlib  Background in astrophysics  Report writing | SQL queries  Writing test suites | I’ve got some extra time on my hands to work on the project  Excited to put all concepts together in a useful project | We don’t have long to complete the project |
| **Saweena** | Report writing  Presentation  Fairly good with python concepts  Chem Eng background- carbon emissions, sustainability knowledge. | Heavy coding | Improving Data analysis skills  Not very good with technical topics but willing to research and learn and get things done. | Low time availability  Need more time to complete heavy coding tasks. |
| **Ashruti** | SQL (beginner level)  Data visualisation with Matplotlib  Process flow design  Visualising content (formatting & graphic design) / presenting  Writing (I’m an English graduate!) | API  Connecting Python to DB | Willingness to learn and contribute where I can  Improving topics already learnt/applying to a real life cases | Low time availability  Background in humanities - unfortunately this won’t help us on the most part but I am willing to try my best to complete this project to a really high level! |
| **Zsoka** | Statistical and logical concepts  ChemEng background  Computing & IT studies (current) | Writing (I’m an Engineering graduate!)  Theory based questions  Giving Presentations | Looking forward to put the learned concepts into practice  Willingness to learn  Fairly new to coding, but finding it straightforward | Low time availability  Upcoming holiday |
| **Yolanta** | Python - uni + currently working as a Data Analyst in the financial services.  Environmental Sciences + Engineering/ Environmental and Climate Change Economics/ Sustainability background  Data Visualisation with Matplotlib  PowerPoint/ Excel | SQL Queries  Connecting Python to DB  Public Speaking (but presenting is fine!) | Looking forward to improving my skills in data science/ data analysis  Willingness to learn and progress  Love working in teams and collaborating with others | Low availability (got a full time job and working Monday - Friday 8am - 6pm which can sometimes go until 6:30pm + 30 min commute) + not available on Sunday mornings.  On holiday 23rd dec - 3rd jan so might be less available/ schedule probably packed and less flexible  Might struggle with heavy coding but will try my best to finish everything on time! |

**Project Timeline**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

