

**Data Analysis – Group Project**

**Group Members:**

|  |  |
| --- | --- |
| **Group Member Name** | **Group Member X Number** |
| Jason O’Connor | X00191019 |
| Brandon Kelly | X00189349 |
| Nathan Kelly | X00182498 |
| Lee Clark | X00189621 |

**Project Overview:**

* **Import Data:** Show and Use Python to Import the Dataset
* **Data Cleaning:** Clean, Pre-Process Dataset with Python
* **Data Manipulation:** Dataset for Analysis & Visualization
* **Data Visualization:** Display Data using Plots & Graphs
* **Hypothesis & Analysis:** Formulate based on Data and do Relevant Statistical Test, Making Interpretations
* **Advanced Analysis:** Include Linear Regression, Multiple Regression, etc.

**Deliverables:**

|  |  |
| --- | --- |
|  | **Deliverables (Submit Via Moodle)** |
| **1.** | Dataset Selection / Collection |
| **2.** | Technical Report: Python Notebook w/ Code |
| **3.** | Report: 2k – 3.5k Words |
| **4.** | Presentation |

|  |  |
| --- | --- |
|  | **Report Overview** |
| 1. | Data Selection / Collection |
| 2. | Technical Report: Python Notebook / Code |
| 3. | Report: 2k – 3.5k Words |
| 4. | Presentation |

**Brief Introduction**

The global onset of the COVID-19 pandemic in 2019 wrought unprecedented changes across the societal and economic fabric, prompting a critical re-evaluation of various sectors. This report centres on one such area of study – the fluctuation in public transport ridership in Ireland, spanning the years 2018 to 2022.

Ireland, as a dynamic European nation, boasts a comprehensive public transportation system that includes buses, trams, and trains. The investigation encapsulates both the pre-pandemic era, characterized by consistent commuter patterns and growing ridership, and the subsequent years marked by lockdowns, social distancing measures, and notable shifts in mobility behaviour.

This report aims to dissect the dynamics of Ireland’s transport systems, addressing essential questions such as: What is the correlation between the decline in passenger numbers and the economic impact on the country? How do the passenger numbers vary across different airports and airlines?

By scrutinizing data trends and exploring contextual factors, our group aims to provide a comprehensive understanding of the intricate relationship between a global health crisis and Ireland's public transport infrastructure. This exploration seeks to offer valuable insights for future national and regional planning and transportation strategies, emphasizing adaptability and resilience in the face of unforeseen challenges.

**Project Management**

Once we had been informed fully about the details of this group project, and we had agreed upon a team, the next logical step was planning and designating work for each member to complete whilst ensuring each member contributed equally and effectively towards the end result of a completed project.

**Group Activity:** Choosing the Dataset.

|  |  |
| --- | --- |
| **Name** | **Tasks Completed** |
| Nathan Kelly | Importing, Cleaning of Data, Statistical Analysis |
| Lee Clarke | Data Visualization, Statistical Analysis |
| Jason O’Connor | Report, Hypothesis Testing, Analysis |
| Brandon Kelly | Statistical Analysis, Seaborn Plot Graphs |

**What Could We Do Better**

During the reflective process of our group assessment, we analysed our efforts through the project, and it has become evident that there are opportunities for improvement that would enhance the quality of our work. With careful reflection, we identified key areas for improvement, which will help us improve the quality of in future projects and the workplace.

* **Enhancing Data Collection Strategies:**

The foundation of this project relies upon the quality of our dataset. Knowing this, we realized there is a massive opportunity for improvement here. In our future endeavours in our next projects, optimizing the search for a good dataset could benefit us by having a more meticulous approach, by exploring more diverse sources, and ensuring the dataset is complete and comprehensive, leading to a better overall experience when working with it.

* **Strengthening Communication Channels:**

We have come to realize that communication is the cornerstone of successful collaboration. Upon reflection, there are many instances of miscommunication resulting in misunderstandings regarding our approaches and resulting interpretations. To mitigate the chance of this happening again, we must implement a structured, organized communication plan, including weekly meetings and check-ins, to make sure each member has a clear vision and understanding of the tasks at hand.

* **Embracing Iterative Feedback Culture:**

When working in a group project setting, we have come to understand that feedback can serve as a very powerful tool for improvement. With a proper foundation in constructive criticism, we can improve our continuous learning. With a mindset of actively seeking feedback on our individual contributions, methodologies, and project management, we can produce higher quality of work leading to a more refined final output.

|  |  |
| --- | --- |
| **Student Name:** Nathan Kelly | **Student Num:** X00182498 |

**Task:** Describing the Data Collection & Cleaning Process, Detailing Challenges & Methods to Counter and Overcome them.

When COVID-19 was first detected in Wuhan, China in late 2019. The world watched on with great uncertainty, especially the aviation sector. The first case in Ireland was detected in February 2020. This middle-aged lady travelled to Northern Italy and returned on 17th February 2020. On the same day, she developed general malaise and cough [1]. This led to panic in Ireland and On March 27, Ireland was placed on full lockdown with all non-essential journeys banned for two weeks [2].

The first travel restrictions were imposed on the 27th of March, where people were not allowed to go on non-essential travel. The aviation sector felt the effect of this. During pre-COVID times (January 2019 - March 2020), the following figures represent the mean of the months during the pre-COVID period.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Pre -Covid Mean (Jan 2019 - Mar 2020)   |  |  | | --- | --- | | Dublin | 2,521,025 passengers | | Cork | 198,856 passengers | | Shannon | 122,176 passengers | | Kerry | 28,373 passengers | | Knock | 60,894 passengers | | Total | 2,931,324 passengers | |

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| During Covid Mean (Apr 2020 - Aug 2021)   |  |  | | --- | --- | | Dublin | 412,326 passengers | | Cork | 15,647 passengers | | Shannon | 14,077 passengers | | Kerry | 5586 passengers | | Knock | 7860 passengers | | Total | 455,496 passengers | |

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Recovery Period Mean (Aug 2021 – Aug 2023)   |  |  | | --- | --- | | Dublin | 2,402,066 passengers | | Cork | 191,986 passengers | | Shannon | 120,463 passengers | | Kerry | 29,874 passengers | | Knock | 60,198 passengers | | Total | 2,804,587 passengers | |

From the results shown in tables, you can see the effect that the COVID restrictions had on passenger numbers. From the period January 2019 - March 2020 the Average passengers in all airports was 2,931,342 per month in the period. When Covid-19 restrictions came into effect you can see a dramatic fall in the average passengers per month. From the period April 2020 - August 2021 there was an 85% decrease in the average passenger numbers from all airports in Ireland. The average during this period was 455,496 passengers. As COVID restrictions started to drop in August 2021 on Non-essential travel, the number of passengers across all airports in Ireland. From the period September 2021 – August 2023 , the average number of passengers that passed through Irish Airports was 2,804,587. This was an increase of 515% on the March 2020 – August 2021 period. The September 2021 – August 2023 monthly average was still 4.3% lower than the period before the Covid-19 pandemic (January 2019 – March 2020).

Standard deviation is a statistic that measures the dispersion of a dataset relative to its [mean](https://www.investopedia.com/terms/m/mean.asp) and is calculated as the square root of the [variance](https://www.investopedia.com/terms/v/variance.asp). The standard deviation is [calculated](https://www.investopedia.com/ask/answers/021115/what-difference-between-standard-deviation-and-z-score.asp) as the square root of variance by determining each data point's deviation relative to the mean[4].

In this dataset I took all the values from all the airports from before the COVID period.

Dublin Cork

|  |
| --- |
| Year VALUE |
| 0 2019 509380.190608 |
| 1 2020 728815.455962 |
| 2 2021 592722.480715 |
| 3 2022 662092.635255 |
| 4 2023 560428.555988 |
| Year VALUE | |
| 0 2019 55042.983692 | |
| 1 2020 57907.816717 | |
| 2 2021 26257.852854 | |
| 3 2022 65477.041749 | |
| 4 2023 61943.069680 | |

Shannon Kerry

|  |
| --- |
| Year VALUE |
| 0 2019 39491.683794 |
| 1 2020 32140.418071 |
| 2 2021 28236.731735 |
| 3 2022 38916.422650 |
| 4 2023 41711.393276 |
| Year VALUE |
| 0 2019 6746.679902 |
| 1 2020 7938.012198 |
| 2 2021 10273.225062 |
| 3 2022 10335.850912 |
| 4 2023 9516.537105 |

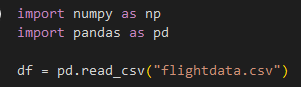
Knock

|  |
| --- |
| Year VALUE |
| 0 2019 16767.094642 |
| 1 2020 15930.348277 |
| 2 2021 16565.455048 |
| 3 2022 22960.149561 |
| 4 2023 23088.000928 |

To get the previous values, I had to go through data manipulation techniques. I started this by creating a Google Colaboratory. Google Colabortory allows anybody to write and execute python code through the browser and is very suitable for machine learning and data analysis.

To help with reading in a dataset, I imported NumPy. NumPy is a Python library used for working with arrays, it also has functions for working in domain of linear algebra and matrices [5]. I also imported Pandas. Pandas is a Python library used for working with datasets, it has functions for analysing cleaning, exploring, and manipulating data [6].

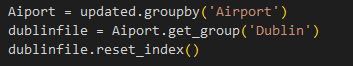
I used Pandas to import our dataset as a Csv file. A csv file is a comma -separated values file. This dataset was the number of passengers that travelled through Irish Airports from January 2019 - August 2023.



As I was looking through the data, I found columns that were not needed. I used the Panda’s feature df.drop() to drop these columns and created a new csv. I used this new csv to split up the airports into different CSV’s.

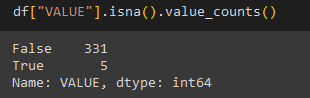


The two panda’s features .groupby() and get\_group() to get this data. The .groupby() allowed me to isolate the “Airport” column whereas the get\_group() function allowed me to isolate the airports that were each airport for example get\_group(‘Dublin’).





By splitting up the main dataset into smaller csv files, it helped me to get the mean and standard deviation of each period. I found values that weren’t numbers which were down as NaN (Not a Number). To help me eliminate the NaN , I used the Pandas function .fillna(). By doing this, it would help me calculate the mean and standard deviation of passengers.

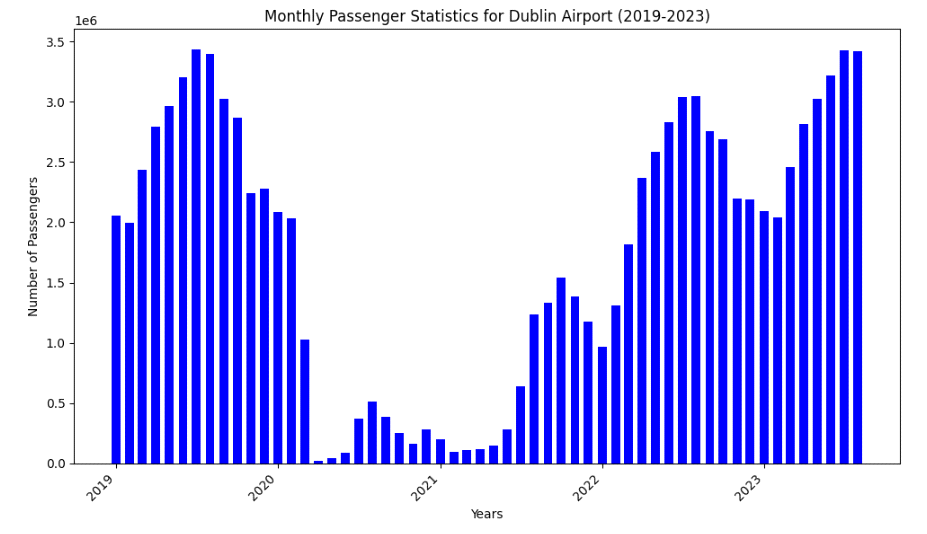
 

|  |  |
| --- | --- |
| **Student Name:** Lee Clarke | **Student Num:** X00189349 |

**Task:** Exploratory Look into the Dataset, Showcasing Data Visualizations & Initial Insights.

The unprecedented nature of the COVID-19 pandemic, which not only had numerous implications for the Irish economy, but the world’s economy, also had an adverse effect on the aviation sector as well, causing it to decline. This decline in the aviation sector also had a hand-in-hand effect on the Irish economy, which is what I will explore further.

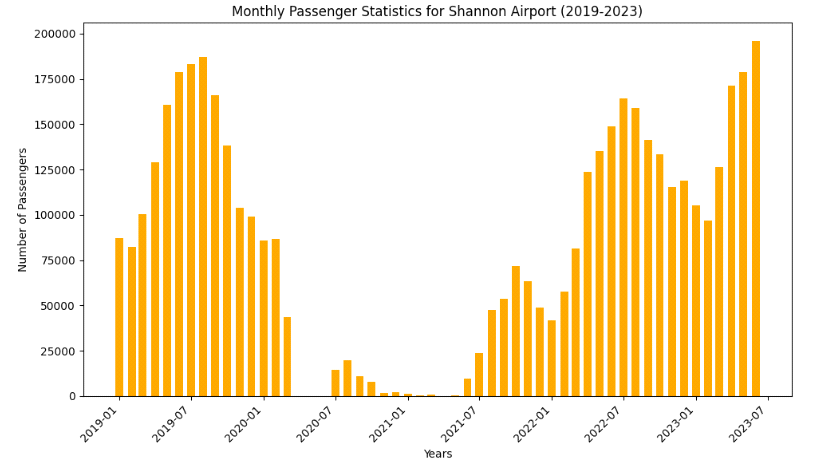
The impact of the pandemic was felt profoundly within the Irish tourism industry. As the virus began spreading globally, there were numerous travel restrictions, lockdowns, and safety concerns which led to a large reduction in both international and domestic tourism. Ireland, largely known for its beautiful landscapes, heritage, and vibrant cities, quickly realised the impact of COVID when there was a sharp decline in visitor numbers coming through the airports, such as Dublin Airport, shown below in the graph.



The graph above, based on Monthly Passengers passing through Dublin Airport, shows clearly in 2020, in the month of March, when the lockdowns were imposed, just how quickly the decline in passengers took effect. Slowly, the numbers began growing again, however, it is clear to see that for the duration of the pandemic during 2020, the number of passengers never grew anywhere close to those of previous years statistically.

The closure of attractions, hotels, and hospitality services led Ireland to a significant financial period of loss for businesses and this caused widespread job disruptions within the tourism sector. The government quickly stepped in, as explained by the KPMG, putting measures into place to reduce the impact and fallout for businesses. For example, as shown by KPMG, just some of the measures included Direct and Indirect Tax Measures, such as deferrals and reductions, whilst also providing Employment-related measures such as state compensation and training schemes.. With the government stepping in to ease the effects of the pandemic, the nature of the impact is clear to see, as it required such a response.

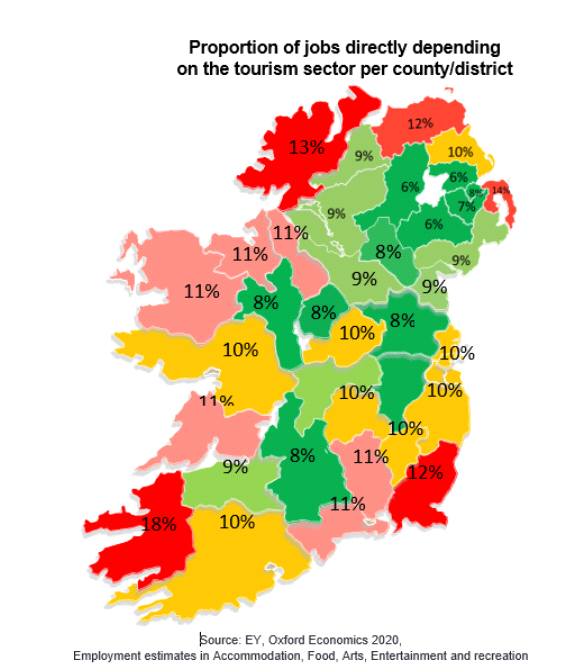
Dublin Airport was not the only airport that saw these declines, as the effect was felt across Knock, and Shannon, meaning Ireland was feeling the full effect of a pandemic, resulting in a massive loss for the tourism industry. Shown in the graph below, we can see that Shannon Airport not only saw a sharp reduction in passengers, but they also had to shut down due to both lack of passengers and demand for the service during the pandemic.



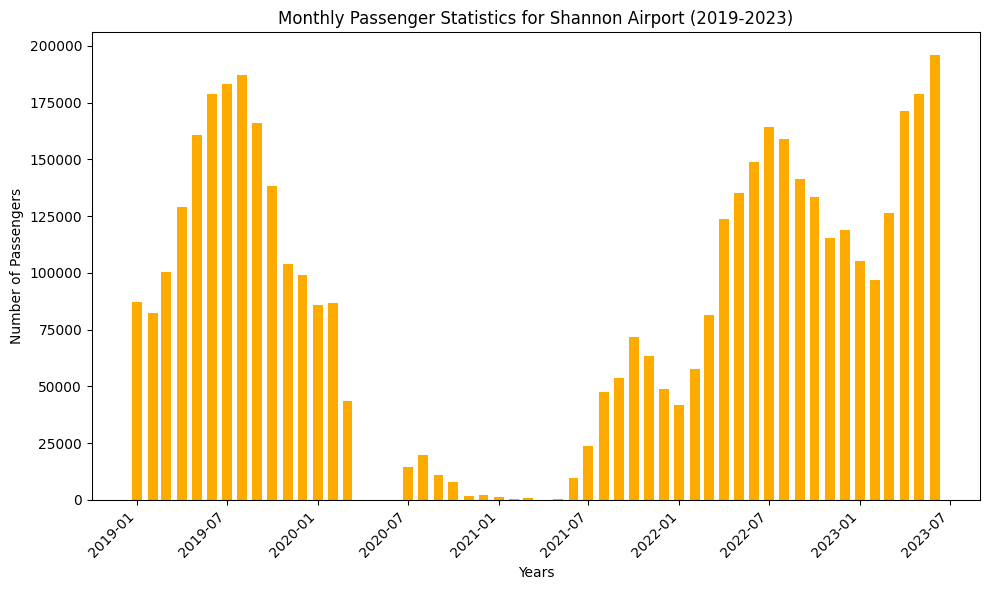
The impact on the tourism industry was felt nationwide, with the website *‘PublicPolicy’* stating that the “Irish tourism and hospitality industry have been hit by COVID-19 restrictions, having lost about €17 million per day, according to the ITIC”. . The article continues to show us that the total losses almost amounted to €12.2 billion, leaving a severe dent in the tourism industry. Not only did the pandemic lead to the closure of airports and monetary losses, but also jobs too. The article states that, of the nearly 260’000 jobs across the tourism sector, over 180’000 were either lost or vulnerable.

As we can see, the closures of the airports, and the loss of international tourists making their way to our Emerald Isle, the impact was not only limited to businesses and their profit margins, but it was felt by the everyday resident of Ireland, with jobs at risk, or lost, due to the lack of visitors to our country. With many of these workers, who were highly skilled in their fields, being left jobless, this caused a massive increase in the recipients of the Pandemic Unemployment Payment, or PUP, leaving them searching for jobs outside their preferred sector, which is a big loss for the country as a whole, as you would wish for specialised workers sticking within their preferred sector.

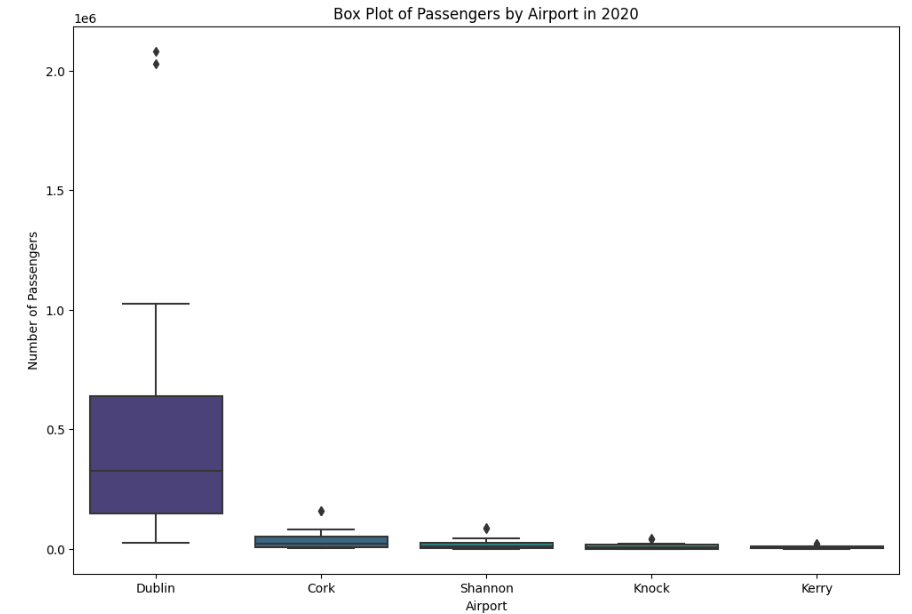
As per the website, *PublicPolicy,* the following graph demonstrates the proportion of jobs depending on the tourism sector per county. Here, we can directly correlate between our graphs, provided by Lee Clarke, and the graph provided by PublicPolicy.



As we can see in the above graph, a county that employees a large amount of tourism sector jobs is Kerry, with a massive 18%. Now, when talking about tourism in Kerry, the closest available airport for the average traveller to reach would be Shannon. However, during the pandemic, the following graph, shows the impact of the pandemic on the Shannon Airport.



As we can see, Shannon didn’t just slow down during the pandemic, it closed completely, which, when we see how many jobs depend on tourism, would have had a massive impact on Kerry, which heavily depends on tourism as a county, with attractions such as Muckross House, Ross Castle, and the Ring of Kerry heavily attracting tourists to the area.



In referance to the above Dublin graph during 2020 ,Just as a comparison between all airports during the year of lockdown and restrictions in the pandemic we can see Dublins numbers still dwarfed that of every other airport in the country, yet it still sufffered heavily from the lockdown with loss of jobs for the airport and shop staff alike.

|  |  |
| --- | --- |
| **Student Name:** Brandon Kelly | **Student Num: X00189621** |

**Task:** What is the percentage decrease in passenger numbers during COVID Period compared to the previous year?

The COVID-19 pandemic triggered an unprecedented upheaval in the realm of air travel, particularly in the domain of airport passenger numbers. As governments worldwide implemented stringent lockdowns, travel restrictions, and safety protocols to combat the virus's spread, airports found themselves grappling with a staggering decline in passenger traffic. This downturn significantly impacted the operational dynamics of airports, leading to financial strain and necessitating fundamental changes in their functioning. The decrease in passengers not only posed economic challenges for airports but also prompted a re-evaluation of their strategies and infrastructure to adapt to the evolving landscape of the aviation industry in the wake of the global health crisis.

In February of 2021, almost a year after the initial lockdown of Ireland the DAA(Dublin Airport Authority) Records €284m Loss Due To Impact Of COVID-19 On Travel[7]. This loss of revenue was mainly due to the drastic decrease in passenger travel in 2021 compared to the previous years.  To understand the significance of this decrease let’s look at the data.

A graph of passengers by airport

Description automatically generated

This bar chart above shows us the number of passengers for each airport across Ireland. Note that the numbers of passenger’s values are in the millions. Upon viewing this chart, you can see that the data shown is as expected before covid.

According to Dept of Transport Ireland [8], the total passengers handled at State and Regional Airports in 2020 were the following figures:

**State Airports**

Dublin:7.3 million

Cork: 527,014

Shannon: 273,585

**Regional Airports**

Knock: 124,532

Kerry: 82,959

Donegal: 18,097

That is a total of 8.3 million passengers passed through Irish airports in 2020, **down** **78.2%** on 2019. With Shannon and Knock experiencing the greatest loss of **83.1%** and **82.3%** respectively. When it comes to state Airports, Kerry, Cork and Dublin all reported a loss of over **77%** in passengers [8].

A graph of passengers by airport

Description automatically generated

As we can see in the Bar chart above, there is a drastic decrease in the number of passengers in 2020 across all airports compared the to the previous year. This was caused by the covid-19 lockdown restricting all travel within the country.

With the dataset we choose, we ran into a couple problems regarding the layout and structure of the data itself. In order this fix these issues we had to manipulate the dataset in a way so that our tests and graphs could interpret them correctly. For example, in the dataset the months and years were combined into one string, this is not ideal as certain packages such as sea born and mat plot lib cannot read these as two separate strings.

To fix this, I used str replace to get rid of any years mentioned within the month column and created a new column called ‘Month Modified’ as not to disturb the original column.

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generatedWith this small fix now complete I can use this modified column to create informative graphs such as this heatmap below

A graph of months and months

Description automatically generated

Note: There are missing data found within 2023 as This particular dataset was recorded before the end of the year.

How do the passenger numbers vary across different airports and airlines? Did some experience a more significant impact than others? With the world at a standstill, airlines had to strategize a way to keep their company afloat during these desperate times. These strategies included a reduction in flight frequencies, flight cancellations, aircraft grounding and cost-cutting measures such as reducing staff hours, or even laying off employees to align with the reduced operational needs.

|  |  |
| --- | --- |
| **Student Name:** Jason O’Connor | **Student Num:** X00191019 |

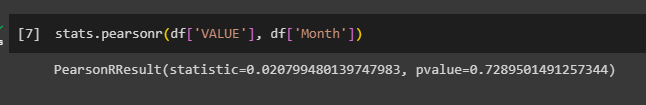
**Task:** Finding out whether there is a correlation between the Month of the Year & the Number of Passengers going through the Airports in Ireland.

To answer this question, the investigation of the correlation between the month of the year and the number of passengers passing through airports in Ireland is a crucial step in the understanding of the patterns and trends in air travel within our country. The dynamic of passenger traffic can be heavily influenced by many factors, including seasonal variations, holidays, and economic activities. During my analysis, I aim to uncover and understand the correlation between these two factors, whilst giving insight into why these reasons exist.

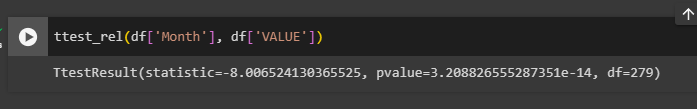
The dataset we have chosen for this group project is based on the years from 2019 until 2023, during which the world was taken by surprise by COVID-19, triggering global responses to combat the spread, which heavily affected airports across the world. In Ireland, once we had reached the stage where lockdowns were effectively required, Knock Airport was closed due to the pandemic, whilst Dublin, Shannon, and Cork were negatively affected by the massive drop in passenger numbers, as detailed clearly by Dublin Airport on their website, where they state *“The number of passengers travelling to and from Dublin Airport and the UK decreased by -24% vs 2019.”*

For the first part of my analysis into the answer to this question, I would be using the Pearsons Correlation Coefficient test. Firstly, I analysed the dataset and decided the two columns I would need to use to correctly answer this is the *VALUE* and *Month* columns, which gives me all the information I would need.

The first test I ran, using the following code, resulted in the output of the P Value of 0.72, which is much higher than 0.05, so I concluded that it was H0, as there is no significant correlation between the two sets of data.



For my second test, I used the Paired T-Test, attempting to answer the same question again, however, this time I was given the answer of 3.208 as the P Value, which at first glance is much higher than 0.05, however, it ends in *e-14,* meaning there is 99 zeros after the final number so it is in fact much smaller than 0.05, giving us the answer of H1 for this testing.



From this testing, it is hard to make a conclusion as to whether there is a definitive answer to this question, as I have received two conflicting answers from the two tests, however, what I can do is analyse this a little further. The most likely answer to be correct, in my opinion, is the Paired T-Test, which showed us that there is a significant correlation between the two factors, as there is evidence to prove this.

When looking at the number of passengers and the month of the year, it is easy to correlate a higher number of passengers towards a seasonal holiday or tradition within the country. For example, Christmas, Halloween, or New Years could heavily influence passenger numbers. As explained by *BreakingNews,* they state:

*“More than 1 million passengers will pass through the airport in December, according to the operator DAA.”* They then continue to state that *“Anyone who is flying in should get to the airport two hours before your flight,”* going to show just how busy the Christmas period is, and concluding in my head that there is a correlation between the months of the year and the number of passengers passing through the airports.

**References:**

‌