Overview:

Slide 1 (Flight Data Analysis):
Hi Everyone! My name is Isaac, and this is group 6

Slide 2 (Outline):
This is the outline of our presentation, where I will start with the overview of our company, Acorn Airlines, followed by efficiency metrics presented by Evangeline, and Punctuality measures by Wenting and Si Xuan. Wu Tong will then cover areas for improvement, and Jing Yun will discuss our proposed course of action.

Slide 3 (Overview):
<Skip Slide>

Slide 4 (Aim):
We analyzed the data to identify areas for improvement to outperform Berry Airlines. Our goal is to provide actionable insights and recommendations that can help us achieve that.

Slide 5 (What is Performance):
So, what do we mean by performance?
First, we need to be utilizing our aircrafts most efficiently. And second, we need to be operating on schedule and arriving at our destination on time. We aim to address the level 1 questions, “How is BA performing better than us?”, “What factors contribute to our arrival delays and how do they compare to BA?”, and “What actions should we take to improve our performance against BA?”

Slide 6 (Proportion of Airlines in Flight Data):
For overall sample size. The square pie chart shows that our flights make up about 47.5% of the total flights, while BA's flights account for 52.5%. This gives us a good baseline for comparison, having roughly equal proportions of data.

Slide 8 (slide 7 hidden, Percentages of Cancellations, Details):
High cancellation rate is another indication of performance issues and our ability to handle unexpected situations. This could impact our overall reputation and customer satisfaction.

Slide 9 (Percentages of Cancellations, Pie Chart):
From these pie charts, the difference between the two airlines is quite small. But as data analysts, we knew we had to dig deeper.

Slide 10 (Percentages of Cancellations, Bar Chart):
Upon closer inspection, we found a significant difference of 9 times the percentage of cancellations compared to BA.

Slide 11 (Key Takeaway):
Based on our analysis, we found to have a higher percentage of cancelled flights compared to BA, suggesting performance issues that require our attention. Now let's hear from Evangeline who will share our analysis on Utilization. 2:06

Utilization:

Slide 12 (Outline):

Moving on I will be talking about Utilization

Slide 13 (Are we efficient):
With level 2 qn of are we efficient?

Slide 14 (Avergae Air Time of Planes):
We explored the data set and made use of the average air time of planes

We found out that on average, our airline AA has 11.05 minutes higher airtime

Than Berry Airlines per plane

Slide 16 (Average Taxi In/Out Time of Planes):

Next, we used the Taxi in/out data set

Taxi in/out is used as indicator of operational efficiency (also part of the turnaround time)

By analyzing the mins we hope that we can identify potential areas for improvements in operations

We found out that our taxi-in/out takes more time than our competitor

This means that we are taking more time on the runway as compared to our competitor,

Slide 17 (Key Takeaway):
The key takeaway is that

Although we have higher average air time per flight, we also incur higher taxi in and taxi and out

It could be an indicator of longer turnaround time of our airline

Hence being less efficient

Slide 18 (Outline):

now we will moving on in analyzing our airline punctuality

Slide 19 (Are we punctual):

This starts with the level 2 qn of are we punctual

Slide 20 (Comparison of Arrival Diverted): Firstly lets start of with the comparison of arrival diverted

On first glance we may not seem to be performing far from Berry Airline

But let's hear from out next speakers

Slide 21 (no heading): Our next speaker will analyze 3a Departure time delay

And secondly our Arrival Time delay

Our next speaker is Wenting 1:25

Departure Delay:

Thank you Evangeline

Slide 22 (Thank you Evangeline):
To analyse our on-time performance, we looked into departure time delay and how punctuality differ for both airlines.

Slide 23 (differ for both airlines):
As shown by this bar chart, we can see we have more flights in the top 20 aircrafts with departure delay, with 14 out of 20 being our aircrafts

Particularly, aircraft N273AY has significantly higher median departure delay time, topping the chart

Overall, our airline is performing worse in departure delay time and **experiencing more delay than BA**
(In this bar chart, we chose to use median because median is barely affected by outliers, giving us more representative results)

Slide 24 (experiencing more delay than BA):

In this chart, it shows that the average departure delay for our airline is higher than that of BA by 12.5%, which is a significant amount.

(However, Since the sample size is large enough for the effect of any outliers being minimal, we also look into average departure delay time for analysis.)

Slide 25 (more detailed breakdown of it):
After looking at the overall departure delay data, now let's look at **a more detailed breakdown of it**.

For us, late flights seems to be a problem since it takes up 41% of the total flights which is almost half the proportion.

Moreover, compared to BA, we have a 6% higher late flights, 2% lower punctual flights and 4% lower early flights

In this case, we are not performing better than BA.

Slide 26 (Next, we try to tackle):
Next, we try to tackle our level two question: what could be the contributing factors towards these results?

With the use of exploratory scatter plot matrix, we see that average departure delay seems to be highly positively correlated with average airline delay and late aircraft delay, thus could be seen as potential contributing factors.

Next we will have Si Xuan to continue the analysis
1:29

Arrival Delay:

[slide 27] (3b arrival time delay)
thank you wenting, now we will be moving on to the second part of our analysis on punctuality, where we analyse our arrival time delay.

[slide 28] (**Top 20 Flights by Median Arrival Delay)**

Firstly, our airline is performing worse in arrival delay time compared to our competitor, BA. We have more flights in the graph of top 20 by arrival delay, with N273AY having a significantly higher median arrival delay time.

[slide 29] (**Airline Average Arrival Delay)**
Next, the average arrival delay for AA is higher than BA by 2.043 mins, which constitutes to a 4.8% difference.
This difference may seem small,

[slide 30] (**Arrival Punctuality)**
however, when we look into the punctuality of the flights, we found that 38% of our flights arrive later than scheduled, which is 6% higher than BA, and our flights are also 6% less likely to arrive early.

Moving forward, we then worked on addressing our second level 1 question, which is
What are the factors contributing to arrival delays for AA and how do they compare against BA?

[slide 31] (**Factors Contributing to Arrival Delay)**
with this, we categorised the factors into controllable and uncontrollable factors.

[slide 32] (**Factors Contributing to Arrival Delay**)
Through our piecharts, we found that the composition of factors contributing to arrival delay is similar for both airlines, with around one third of delays caused to factors beyond the airline's control.

[slide 33] (**Factors Contributing to Arrival Delay**)
This brings us to a level 2 question which is : What are the significance of each of these factors towards arrival delay?
As seen, both airline have more delays caused by controllable factors and the most significant factor for both airlines is airline delay. Our analysis also concludes that we should target these controllable factors, especially aircraft delay.

[slide 34] (Key takeaway)
Finally, our key takeaway is that compared to BA, our airline has higher median delay times, lower punctuality, and significantly more late aircraft delays.
I will now hand over to Wutong, who will discuss the identified areas for improvement
1:40

Area for Improvements

Slide 35(outline)
[Transition] Thank you Si Xuan. The analysis on areas of improvements aims to answer the last level 1 qn: What actions should we take?

Slide 36(skipp)
[can skip through]

Slide 37(Key Factors affecting our performance)
There are three factors dragging our performance: a much higher cancellation rate, more serious arrival and departure delay, and longer taxi in and out time.

Slide 38(Breakdown of Cancellations)
For cancellation rates, weather and airline or carrier issues are the most prominent factors, causing 55.62 per cent and 36.76 per cent of cancellations respectively. Therefore, these two factors should be investigated further to reduce cancellation rates.

Slide 39(Departure and Arrival Delays)
Next, N273AY is the most problematic plane with a notably higher median overall delay time. Hence, it is another subject of close examination and further modification.

Slide 40(Taxi In Time)
For taxi in time, we have determined the top 10 airports with remarkbly higher average taxi in time than BA to pinpoint the airports to improve the efficiency of the process

Slide 41(no heading)
For these airports, we look at the top three planes with the longest taxi in time. They should be put into further investigation and subject to improvements.

Slide 42(Taxi out time)
The same analysis goes for taxi-out time. We also discover two recurring airports here, PHX and PHL, and these two airports should be paid extra attention to.

Slide 43(no heading)
Similarly, this graph provide another list of planes to focus on for taxi-out time. <- 10min 20s mark

Slide 44(Outline)
Next, I will pass my time to Jing yun.1:34

Moving forward:

Slide 45 (Moving forward):
Thank you Wu Tong. So moving forward, how should we work to better our airline?

Slide 46 (Solutions to reduce cancellations):
Firstly, to reduce cancellations, we should perform a further investigation to identify the root cause of our high cancellation rate. Some possible areas to look into include maintenance and crew availability.

Of course, there are a lot of cancellation factors that are not really within our control, such as bad weather and airport delays. For these factors, we can only look to using predictive analytics to best predict future trends and to come up with preventive measures against them. Such measures may include improving scheduling and developing alternative routes.

Slide 47 (Solutions for delays):
As for our delays, our first course of action should be to identify the root cause of our delays. Particularly, we should look into the main reason causing flight N273AY delay, since it has a notoriously high arrival and departure delay. We should also look into what is causing our airline delays and late aircraft delays.

Then, some possible improvements we can make include firstly, optimising our flight scheduling and the efficiency of our pre-flight processes. Also, we should look to streamlining our maintenance processes so that our flights can depart on time. Additionally, it never hurts to be prepared and to plan for contingencies and have backup resources to minimise the impact of delays. Lastly, we should figure out the problems with N273AY and see if it applies on a wider scale.

Slide 48 (Future recommendations):
In the future, we recommend to conduct predictive analytics to determine how proposed solutions may impact delay, and to conduct prescriptive analytics to determine the most optimised way to move forward

Slide 49:
We hope that our insights will be useful in bringing our airline to greater heights. Thank you all for listening to our presentation. 1:43

Possible questions

1. Having you done any types of data cleaning? Why or why not

(DC: what to do with null values or outliers or editing any correct or missing data, or reformatting data, etc)

For analysis such as **average delay time**, we excluded the non-positive values because we are primarily focusing on the case of being late then scheduled. This is especially for arrival delay, because late arrival can have a ripple effect on other aircrafts delay.

For **factors contributing to arrival delay**, we also reformat the data in order to use count() function in tableau. Previously, the data was null for flights with no delay, and the value in each cell is the the duration of the delay, hence we could not get the desired graph. Therefore, we reformat the data such that every cell is either 0 or 1, with 0 suggesting the absence and 1 suggesting the presence of delay.

For others, we include all data to obtain a full picture of the entire sample data.

2. Why did your group define performance as such?

- We felt that optimum aircraft utilisation and on time performance are two of the most important determinants of an airline's efficiency, since there are alot of costs associated with not maximising the use of aircrafts. This includes a whole bunch of maintenance costs and taxes that can really set back the overall performance and profits of the airline. Additionally, on time performance is crucial for maintaining the carefully planned out schedules and for overall customer satisfaction, which is again tied with the overall efficiency of the airline and the attractiveness of our service to potential customers.

3. What was your motivation behind analyzing a specific factor?
    a. Why did you use taxi in out time as a measure of operational efficiency?
        i. Since we defined an airline's performance using the factors of optimum aircraft utilisation and on time performance, to ensure that our airline is using the planes as optimally as possible, we should ideally try to minimise taxi in and out times. This is because shorter taxi times will mean shorter turnaround times, which means that our planes can spend a longer time in air and therefore be utilised more efficiently. Additionally, longer taxi times, or taxi times that are longer than expected can lead to delays and negatively impact our airlines on time performance. By analysing these metrics, we can then identify potential areas for improvement in our operations, such as optimising gate assigmnments, improving communication with ground crews, or implementing more efficient taxiing procedures.

4. Why did you not analyse ___?
    a. Turnaround time(the time that passes from landing until take off for a new flight.)

i. Since delay and cancellation are important issues we face currently, we feel that we should address the problem of delay, especially arrival delay first before scheduling flights closer to one another. Else, our cancellation rates and delay time might go up further.

b. Diversion

i. As shown in our pie chart, there is no significant difference between the proportion of flights diverted of both airlines–which means that neither BA or AA is performing worse or better in terms of minimising flight diversions. Since our aim is mainly to look at ways to improve ourselves against BA, we thought that this area would be less pertinent in our overall analysis.

5. What are some possible confounders?

A confounder (or 'confounding factor') is something, other than the thing being studied, that could be causing the results seen in a study.

6. Why did you choose this method of storytelling?

- The aim of our analysis is to better our airline and improve ourselves against our competitors, such as berry airlines. Hence, our storytelling method heavily focuses on comparing between the different factors of the two airlines, so that we can see the ways in which our airline is lacking behind Berry airlines. That way, we know that these are the areas our airlines should work on to improve in the future. We also opted for a broad-based analysis so that we can figure out the areas that our airlines is most lacking in and that are therefore most pertinent to improving our performance. This will allow us to constantly strive to better our services for our customers.

(senior's answer) We decided to use a mix of analysis and vision for the future in our storytelling, because we felt that these 2 methods complement each other, especially in the context of the population health programme, where the overarching aim is to increase population health and not just to increase the health screening process. By using a mix of these 2 methods, we are then able to look at the present and the future. For the Analysis part, we used the given data to find patterns in attendance and demographic, understanding reasons for low attendance and follow up rates. By finding patterns in these datasets, we were able to use it to analyse for reasons which springboarded our suggestions, which then leads us to the next portion of our storytelling, which is the vision for the future. We wanted to focus on how to extend PHP beyond current scope so we used it to propose solutions in which we can improve and this came together to form our overall story.

7. What was the motivation of choosing the top3/10/20 (i.e. the number)?

Sample size high enough to eliminate outliers,

Top 20: trend eventually

1. Improved Visualization: By displaying only the top 20 results, you can provide a clearer and more concise representation of the data. This will make it easier for your audience to understand the information and identify any trends or patterns.
2. Enhanced Focus: Displaying too many results can often lead to clutter and confusion. By limiting the number of results, you can help your audience focus on the most important information and avoid overwhelming them with too much data.
3. Consistent Trend: As you have mentioned, the trend eventually evens out. By displaying the top 20 results, you are providing your audience with a consistent and reliable trend that can be easily understood and analyzed.

Overall, by presenting only the top 20 results, you can provide a more effective and efficient visualization of the data, while also ensuring that your audience can easily understand the trends and patterns being presented.

Scatter plot

- 1. identify any patterns or relationships between the two variables being compared.
- Factors contributing to flight delay
- Different gradients, abs value closer to 1, the stronger the relationship
- A scatter plot is a type of graph that displays the relationship between two variables. The purpose of using a scatter plot is to identify any patterns or relationships between the two variables being compared.
- For example, if the dots form a straight line from bottom left to top right, this indicates a positive correlation between the two variables. On the other hand, if the dots are scattered randomly across the graph, this suggests that there is no clear relationship between the two variables.
- 2. Detect outliers:
- 3 hundred thousand Scatter plots can also help identify outliers, or data points that are far from the general pattern of the data. Outliers can be important because they may
- 3. Sample size is large, : Scatter plots can be used to visualize large datasets, allowing us to easily see how the data is distributed and whether there are any clusters or patterns.