Student: Gourav Verma
Class: DSC 640
Professor: Benjamin Schneider

Fictious Airways Presentation

For my presentation, I chose to use the story telling strategy of the three-act structure. Starting with designing the story boards to use the first part of building the stage, the characters involved. Second to build out the problem, the conflict, and the tension. And third, the solution to the problem. I believe this story telling structure worked well for the story behind my video presentation. It allowed for my slides to transition nicely.

The slides of my video presentation contain 18 slides. In my first few slides, this strategy allowed me to start my story by introducing our plot and protagonist – the reputable world class travel company named, "Fictious Airways" and define a relatable environment of air travel. On the second slide, to supportively portray a world class travel company, I was able to use a spatial map along with sized bubble marks to represent travel locations and the size of travel popularity to these travel destinations. Air travel as we all know, is a common form of transportation used for vacations travel, business trips, travel to major universities, and many more.

In my next few slides - between slides 3 to 5, I was able to introduce the problem that our protagonist was faced with regarding the widespread concern of air travel safety made public by news reports showing increased counts of airplane fatalities. The supportive data visualization used in these slides include a stacked bar chart that displayed the year over year differences by months between March to September. It measures flight counts for each month comparing last year with the current. In this graph I was able to use blue to portray last year's counts (before the news of flight concern was released) and contrast this using burgundy to represent the current year. This graph helped define and support the conflict with my story relayed by the dramatic dip in travel that was reflective of the flight concerns that would lead to potential bankruptcy.

The rest of my slides between slides 6 to 18, were structured to present the third act. In this phase, our protagonist constructs the solution to the flight concern problem. They start off by hiring a new team of data scientists. A data science team with a mission to analyze the data of the environment and provide actionable choices to help improve customer confidence in air travel with Fictious Airways. The team of data scientist then set off to perform a series of data analytics to answer multiple questions. Each of these questions are defined on slides 7 to 17.

On slide 7, they tackle the analytic breakdown of the recorded 297 aircraft crashes using a donut chart with arcs structured to compare crash counts by type of aircraft. One of the takeaways from this donut chart was the significantly high number of crashes for the 737-200 type plane with 115 crashes.

On slide 8, they tackle the question if age of the aircraft contributes toward the crash fatalities using a funnel plot using bar for each of the types of aircraft and colors to indicate which had a survival rate of less than 30% or greater than 60%. This graph helped identify that one of the oldest planes does have less than a 30% survival rate however, it also proved that age is not the contributing factor as some of the other less than 30% survival rate planes were not the oldest aircraft.

On slide 9, they review if the location of the crash affects the number of fatalities using a spatial map with sized pie charts relative to the amounts of fatalities per the location. And pie sections colored to match the type of aircraft involved in the crash. This graph also displays the legend in a sorted manner to indicate the planes with the most crash fatalities with purple.

On slide 10, they reviewed the relationships using a heat map comparing the aircraft types with several metrics including number of crew members, crew fatalities, passengers, passenger fatalities, aircraft age, total fatalities, if the crash had survivors, crash counts by plane. With this graph the were able to identify that the 737-200 that earlier shown to have the most crashes also had the most passengers on board.

On slide 11, they wanted to emphasize the story identified by the heat map on slide 10 regarding the trend of the higher than 60% survival rate per the Boeing 737-200 aircraft. To intuitively indicate that even though this aircraft was involved in the most crashes – it still has a survival rate above 60%. To do this, the data science team rendered a horizontal bar chart that compared the amount of survivor's against fatalities counts.

On slide 12, the wanted to expand on the differences of magnitude shown on the horizontal bar charts on slide 11. The then decided to change the horizontal bar chart visualization to a stacked bar chart to further make the the survival counts and fatality counts more obvious for each of the different plane types. By doing this, it made it easier to see which aircraft had fatality counts larger than survivor counts therefore raising the question about three specific aircraft types to be reviewed more in the following slides.

On slide 13, they wanted to visualize the importance of the "survival rate" metric. Here they used a bullet chart to convey the analysis between the survival rate per each aircraft with gauging sections highlighting which aircraft had a score lower than 30% and those with higher than 60%. They used blue to define the good scored section. And pink as the in between good and bad color. And dark red as the bad scoring area. This graph helped emphasize the review of the 3 aircraft types that have less than 30% survival rates.

On slide 14, they wanted to update the executive team of Fictious Airways however, they knew that that the members of the executive are less inclined to understand graphs. So, they created a simplified data grid containing the current survival rate score for each aircraft type and the aggregated survival rate for the company.

On slide 15, they wanted to expand on the story provided by the table presented on slide 14 to highlight the need to decommission the 3 aircraft types identified earlier in the bullet chart that have survival rates lower than 30%. To emphasize the relationship of this data grid to the

bullet chart on slide 13. They used the same colors used to mark the scoring sections of the bullet chart.

On slide 16, they wanted to display supportive analytics regarding the 3 aircraft recommended to be decommissioned. To do this they shared stacked area charts of the 3 aircraft over time to detail the counts of aircraft passengers and plane crew. Compared with a separated set of stacked area charts that displayed the counts of fatalities per aircraft over time for crew fatalities and for passengers.

On slide 17, to share the forecast with the executive members of Fictious Airways, the data science team shares an updated table grid displaying the remaining aircraft excluding the 3 decommissioned aircraft. And the new improved survival rate score of 67%.

On slide 18, the presentation closes using an emphasis on the improved survival rate score for Fictious Airways after having decommissioned the 3 aircraft that had a high likeliness of not surviving an airline crash. Therefore, improving the likeliness that flying with Fictious Airways will more than likely increase your chances of survival if involved in a crash.

How the data was prepared.

To prepare the data and visualizations used for this video presentation, I created Python scripts to dynamically scrape the 26970 URL pages from the Bureau of Aircraft Accidents Archives web site - listed in the data source section. After scraping the data for each URL, my Python scripts then loaded the data to a relational database on MS SQL Server. All data wrangling steps were completed on the MS SQL relation database and within PowerBi. Within the data wrangling phases, I cleaned and removed the fields: YOM (year of manufacture using the majority of rows that had the same model plane), flight date corrected the format as some rows were different, and operator as some rows contained a unique identifier in the names enclosed in parenthesis. In addition to cleaning up rows, I used SQL to create the following fields: PlaneAge derived from the YOM and the flightdate. As well as performed exploratory data analysis to better understand the distribution of my data. I then generated csv datasets from my relational db to consume within PowerBi and Tableau. Using PowerBi, I added the following fields: [Number on Board] derived from the [Crew on Board] and [Pax on Board], [Survivor Count] derived from [Number on Board] minus the [Total Fatalities], [Survival Rate] derived from [Survivor Count] divided by [Number on Board]. Since I found that I needed latitudes and longitudes per country to enable a map presentation of the data. I imported a country lat-long data file and used the Power Bi DAX language to create a formula for the country name to allowed me to join on the country name fields.

Differences between presenting the data story to a general audience as opposed to an internal audience.

For this project, my goal was to create a presentation that was targeted for a general audience to share at a level that is relatable to any user who is familiar with air travel. Versus targeting an audience who is aware of the specific terms that are specialized and mostly

known by those who work in the air travel industry. I believe I was able to do this by using familiar generic air travel terms, graphics that are simple and organized to prevent confusion.

How would I do this differently if had to do it all again from the beginning.

If I were to do this project all over again, I believe I would have created a more animated video using a tool called Adobe After Effects. The version of Adobe Creative Suite does not work on my Windows 10 workstation and thus did not have the option to use the tool.

Data sources:

- [Accidents and Fatalities Pery Year](#)
  https://docs.google.com/spreadsheets/d/1SDp7p1y6m7N5xD5_fpOkYOrJvd68V7iy6etXy2cetb8/edit#gid =1448957446

  Supplemental data:
  [Bureau of Aircraft Accidents Archives](#)

  U.S. Department of Transportation
  https://data.bts.gov/
  https://data.bts.gov/browse?category=Aviation
  https://data.bts.gov/Aviation/Commercial-Aviation-Departures/bpqk-hyst

  Bureau of Aircraft Accidents Archives
  http://www.baaa-acro.com/crash-archives
  http://www.baaa-acro.com/statistics

- [Crash Datasets](#), Data.World
  Airplane Crashes 1908-2009
  DATASET BY HOZEFA HAVELIWALA
  https://data.world/hhaveliw/airplane-crashes-1908-2009

  To get country lat and long addresses, I also used: countries.csv.
  https://developers.google.com/public-data/docs/canonical/countries_csv