Title slide, introduction

Professional Summary

Background in special education - Special education different than what you think of teachers. Nowadays everything is about documentation, paperwork, and data-driven decisions which is what got me into data analysis.

Work at Auburn University as an administrative support assistant - I do mostly analysis and reporting, designing data presentations, and organizing events. I use my excel, sql, and tableau skills i have learned in this class every day at work.

Desire to transition to tech industry - My next steps are to get a masters degree in information systems and go from there! Still trying to figure out the best avenue for this. But I am excited and this bootcamp has given me a lot of confidence in my abilities.

Python Analysis

This is my python analysis. I am breaking everything down into steps. First step is the problem.

A production company wants to analyze TMBD data to predict whether their movie will be successful. Go to approach slide.

Go to jupiter notebook and display your notebook. Go through the dataset -

The dataset contains 4800 rows and 20 columns.

highlight key columns - budget, genres, popularity (higher the number, the more popular), production companies, revenue, runtime, title

First I identified the null values and performed null value treatment to clean up the data.

Python Analysis

It is important to clean your data before analysis to make sure important data is not missing or there inconsistencies. Its better to clean before you start to save time and get acquainted with your dataset.

First I identified null values by using is null. Go through process on jupiter notebook.

Every important column has a value and I am ready to begin analysis.

Analysis - finding movies with a budget above 220k - I created a new dataframe to contain all the movies with a budget greater than 220k, I then imported json to get all of the genres in one column and be able to tally them, I did this by iterating over rows in a dataframe

Python Analysis

I created another dataframe to be able to group the genres - these are all the genres associated with movies above 220k budget. Which really is not a high budget at all. Almost 80% of the movies in the dataset have a budget over 220k.

The most frequently occurring genres are Drama, Comedy, Thriller, Action and Adventure.

I did the same thing for movies that have a revenue above 961M. This revenue is the top 25 percent of revenues, greater than the average. I added their genres to a dataframe and counted the occurrence. The most frequently occurring genres for movies with a high revenue are Adventure, Action, Fantasy, Science Fiction, and Family.

Python Analysis

Some movies had 0s as values for their budget and revenues. I created a new dataframe where none of the values were 0 for budget or revenue.

I wanted to know the top 10 highest revenues. Top 10 highest revenues are Avatar, Titanic, The Avengers, Jurassic World, Fast and Furious 7, Avengers: Age of Ultron, Frozen, Iron Man 3, Minions, and Captain America: Civil War. Go to slide to show breakdown of genres

Lowest budget movies are: Modern Times, A Farewell to Arms, Split Second, Bran Nue Dae, The Prophecy, Of Horses and Men, Nurse 3-D, FIST, Angela's Ashes, and The 51st State. None of these movies had too high of a revenue.

Briefly explain the lowest budget top 10 and how it is invalid.

Python Analysis

I wanted to see if there was a correlation between popularity and budget. Go back to slides.

Minions had a budget of 74 million and the highest popularity of 876. The least popular movie of the top 10 revenue was Iron Man 3 with a score of 78 and a budget of 200 million. Mean is 29M

Iron Man 3 earned more revenue than minions, but minions was more popular and had a lower budget.

Python Analysis

I wanted to know the most popular production companies. There were over 13,000 production companies. The most common 5 production companies are Warner Bros, Universal Pictures, Paramount Pictures, Twentieth Century Fox Film, and Colombia Pictures.

These are top 5 based on the number of movies produced.

Show slideshow about the top 5 production companies, and then the overall popularity vs budget

Python Analysis

Go back to Notebook: Outliers: This is a graph with zeros in budget. There are a lot of outliers in budget. A lot of distribution. Median is 15M, Mean is 29M. In the box plot without zeros, the median 25M and the mean is 40M. Excluding zeros did not really change the distribution a whole lot because we still have a ton of big outliers. The mean is still greater than the median which means that there are a lot of lower values compared to the higher ones.

Lots of outliers in Revenue this is with zeros. Median is 19M, Mean is 82M. Without zeros, the median is 55M and the mean is 121M. Right skew, meaning that the median is closer to the middle of the data because there are a lot of lower values.

Runtime outliers: Median 103, Mean 106.

Python Analysis

Overall, there is a bit of a correlation between popularity and budget. 0.4 which is relatively strong.

There is no correlation between popularity and runtime.

My recommendations are to go with a big production company that has a lot of resources for marketing and budget. I think that's what Disney and Warner Brothers have, and their movies do well because of it. Bigger budget means more popularity most of the time.

Appeal to large population groups with genres that include many different types of people like Family, Action and Adventure.

Tableau Analysis

For the tableau analysis I analyzed another dataset from a ficitious bank. They want to do a customer churn analysis which involves me looking at who exactly is leaving the company and find ways to reduce customer loss rate.

Go to approach slide.

Go to Jupiter notebook. Preview the dataset and explain columns.

Go back to slides and do outlier analysis.

Go to Tableau and go over the dashboard.

Tableau Analysis

Most of the attrited customers are female, from England, 40-45 years old, have graduate degrees and are either single or married.

Go back to slides and present final information.

Average credit card transactions of attrited customers is 45. Average of existing customers is 69. How can we encourage customers to use the card more?

Average times the bank has communicated with attrited customers is 3. Average existing customers is 4.

Tableau Analysis

Overall, we need more customer surveys and input. What are customers liking about their credit cards? The customers who do stay, why do they stay? We need to figure out why almost 20% of our customers are not using their cards. We need to personalize marketing and service strategies for each demographic.

We need more customer communication. In the last year on average, we only interacted with active customers 4 times! And inactive customers 3 times. We need to increase communication to better serve our customers.

Tableau Analysis

92 percent of customers are still in the blue card category, the lowest tier. We need loyalty programs to get the majority of long-term customers in higher tiers with better benefits. We can partner with other businesses like hotels or airlines and attract people who travel frequently.

We need to incentivise them to use their card. This could be atm rebates or better cash back percentage. Incentivise them to spend more money with benefits based on spending levels.

Stay informed about competitors' offerings and market trends. Adjust the bank's credit card features, interest rates, and benefits to remain competitive in the industry and retain customers who may be enticed by rival offerings.

Opportunities for Improvement

- Python: actors - there is something to be said for movies that I watch just because there is an actor I like in them. It would be useful to determine the most popular actors for movies that would be sure to make your movie a hit.
- Do ratings of movies affect the popularity of a movie? Does a G rated movie do better than an R rated one and why?
- Dataset accuracy - researching actual budgets online and they were different than the data points in the dataset.
- I would also like to see a dataset from different eras. There are so many movies from the 40s and 50s that it would be interesting to analyze them on their own without worrying about inflation and how budgets and revenues compare to today’s inflation.

Opportunities for Improvement

- Tableau: knowing why customers left. It would be helpful to know why the attrited customers left so the bank can figure out what they need to do to increase their customers.
- Dataset accuracy: actual dependent counts and too many unknowns in credit limit, education, marital status, etc.
- I want to know how long this bank has been in business and more about them to help them with customer satisfaction and incentives. What kinds of incentives do they do already and what is not working for them?