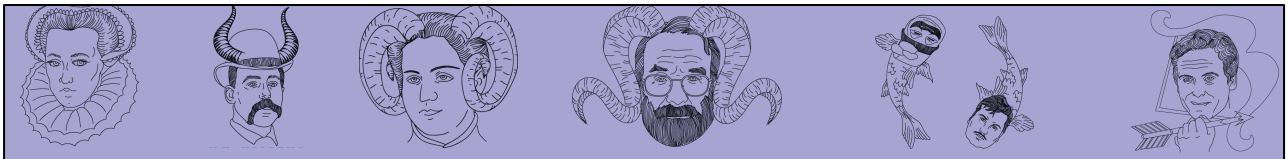
HORRORSCOPES

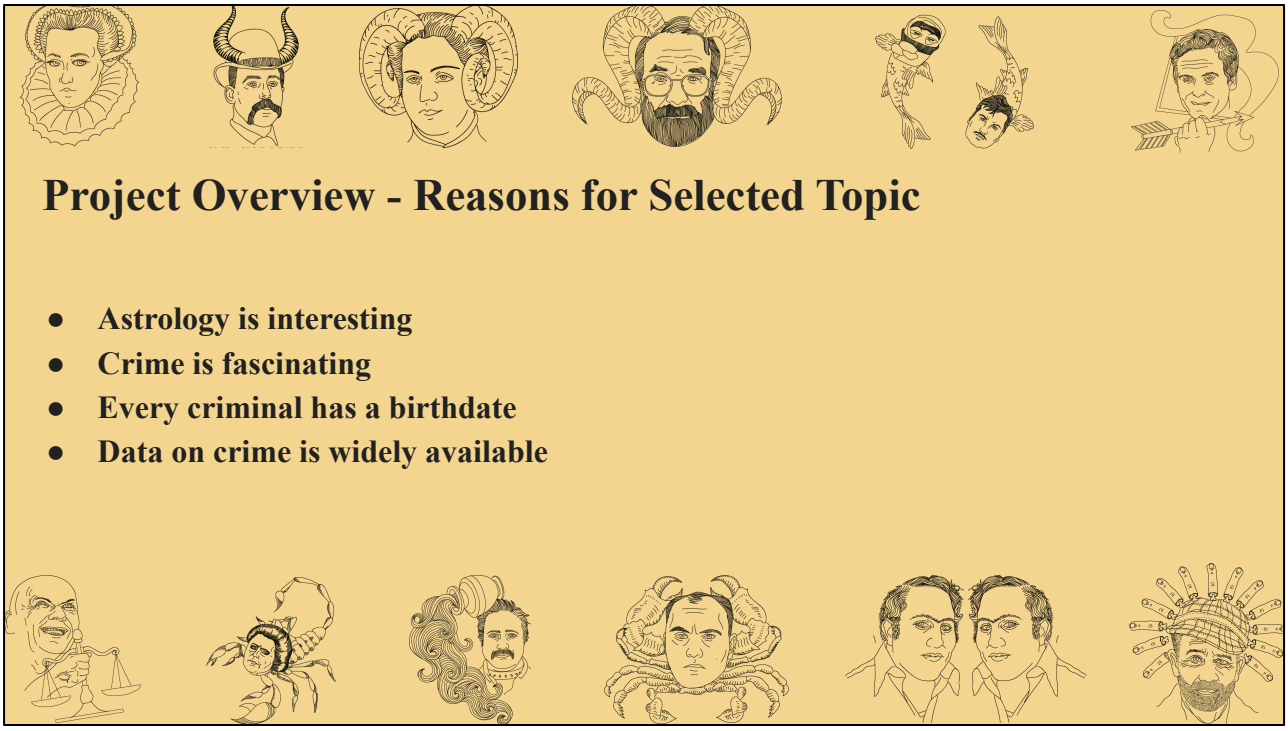
Does astrology influence crime?



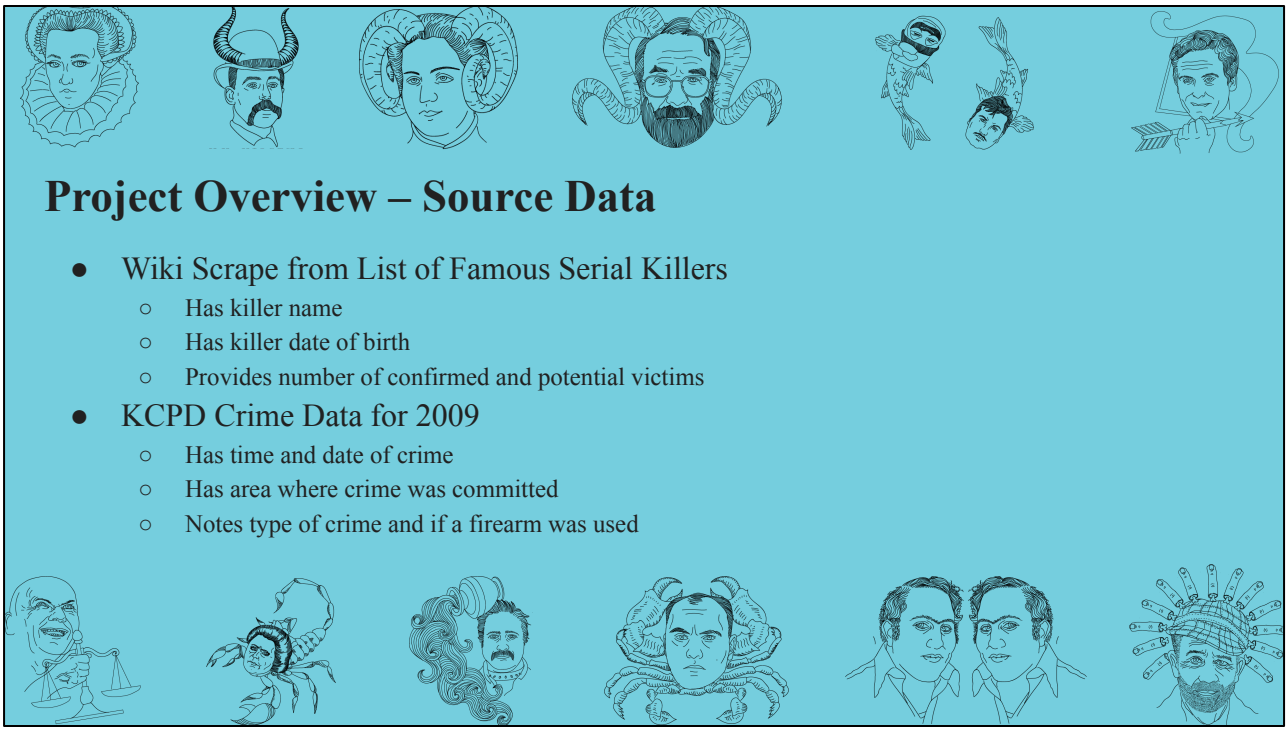
Project Overview - Topic Selection

Our group chose the topic of astrology and its influence on crime for our final project. We all have a shared fascination with both astrology and crime, and were curious as to whether it could be proven that there is a correlation between the two.

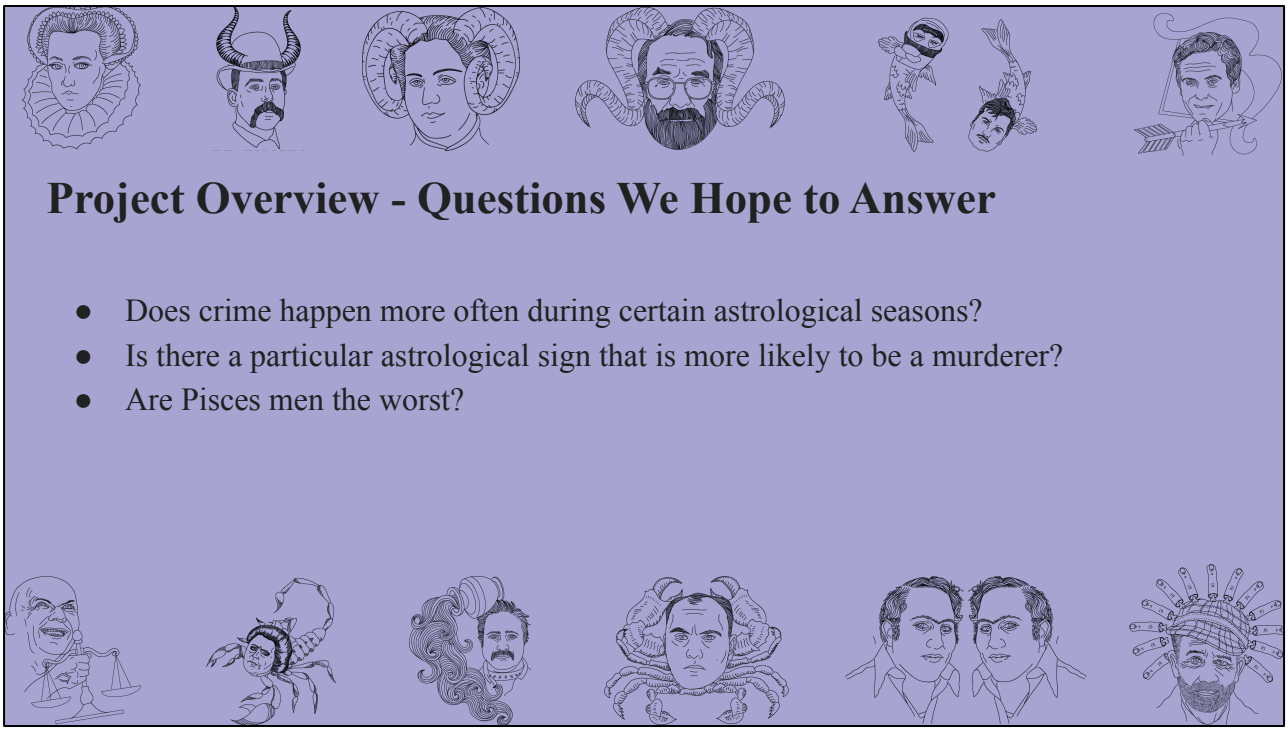
Darcy: Discuss and acknowledge the whimsical nature of astrology, and how people often search for an explanation as to why crime happens.



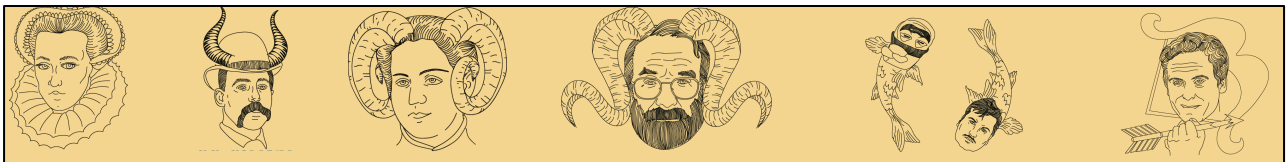
Darcy: Talk about how we found our data and how we determined that the initial data was suitable for our purposes.



Darcy: Discuss how we scraped data from wikipedia and how KCPD data was already formatted into a CSV.



Darcy: Briefly discuss hypothesis that Pisces Men would be the most likely serial killers.



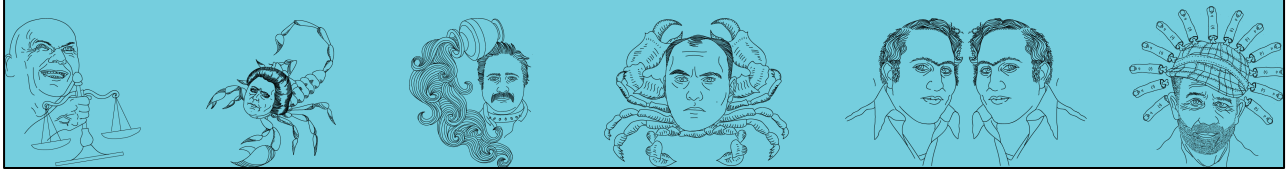
Project Overview - Data Exploration - Wikipedia Scrape

- Once originally scraped, we found erroneous entries for Wikipedia articles such as “The Dating Game” and “September 11th attacks” as well as duplicate entries for several killers
- When that error was fixed, we then encountered a handful of entries that only noted the year the killer was apprehended – these entries were removed.
- Code was then written to assign an astrological sign to each killer by their date of birth
- Data was then exported to a CSV

Darcy: Note that we scraped from a table on Wikipedia. Discuss finding the erroneous entries and troubleshooting that was done to resolve the issue.

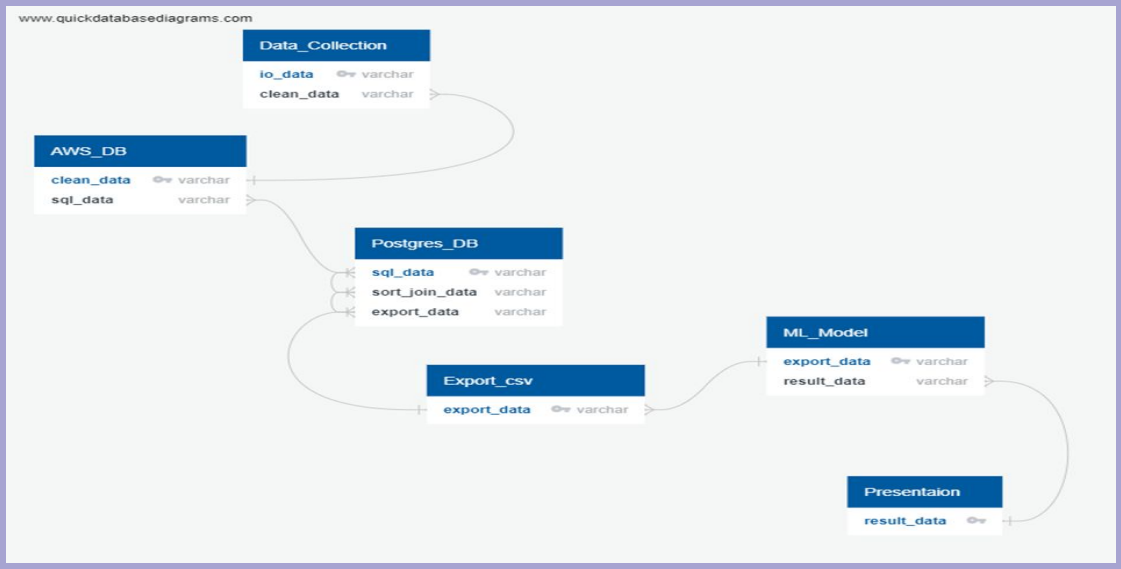
Project Overview - Data Exploration - KCPD Crime Data

- We removed unnecessary columns, leaving only the report number of each crime, the "from date", the "from time", description of the crime, whether a firearm was used and the classification of the crime.
- Data ranges from 01/01/2009 to 12/31/2009 and has over 100,000 rows.
- Once cleaned, data was loaded to our database.



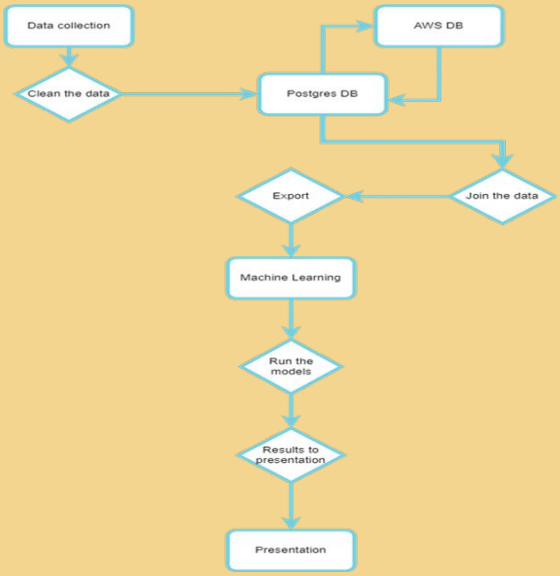
Darcy: Discuss cleaning of this data and potentially how we had to format the date column.

ETL - Schema

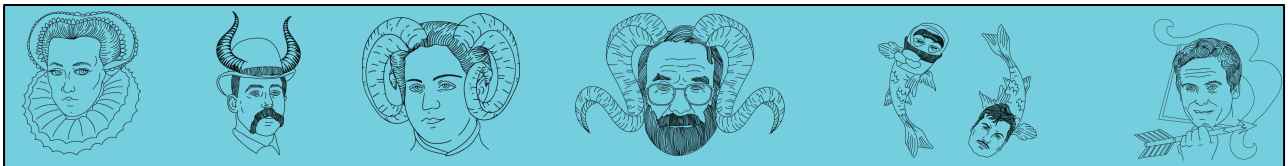


Angelique: Speak to the schema and what connections there are

ETL - Outline

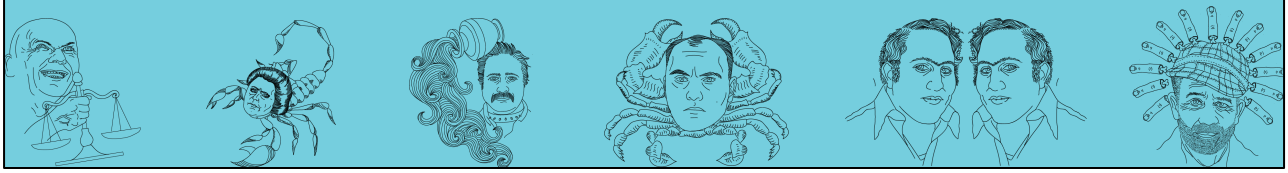


Angelique: Speak to our process here, explain the outline

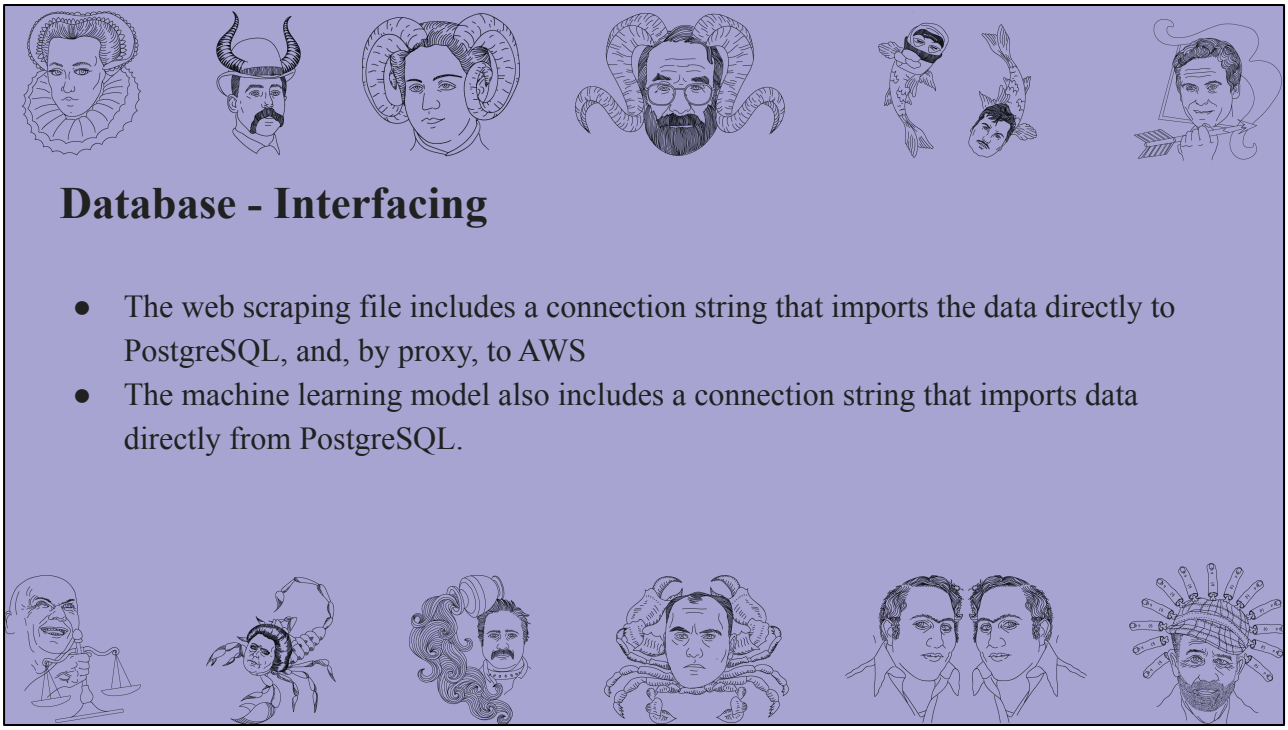


Database - Description of Database

- Hosted instance on Amazon Web Services, access through secure link with pgAdmin through a PostgreSQL database.
- PostgreSQL allows the entire team access to the database on their own terminal, reducing the need for data transfers.

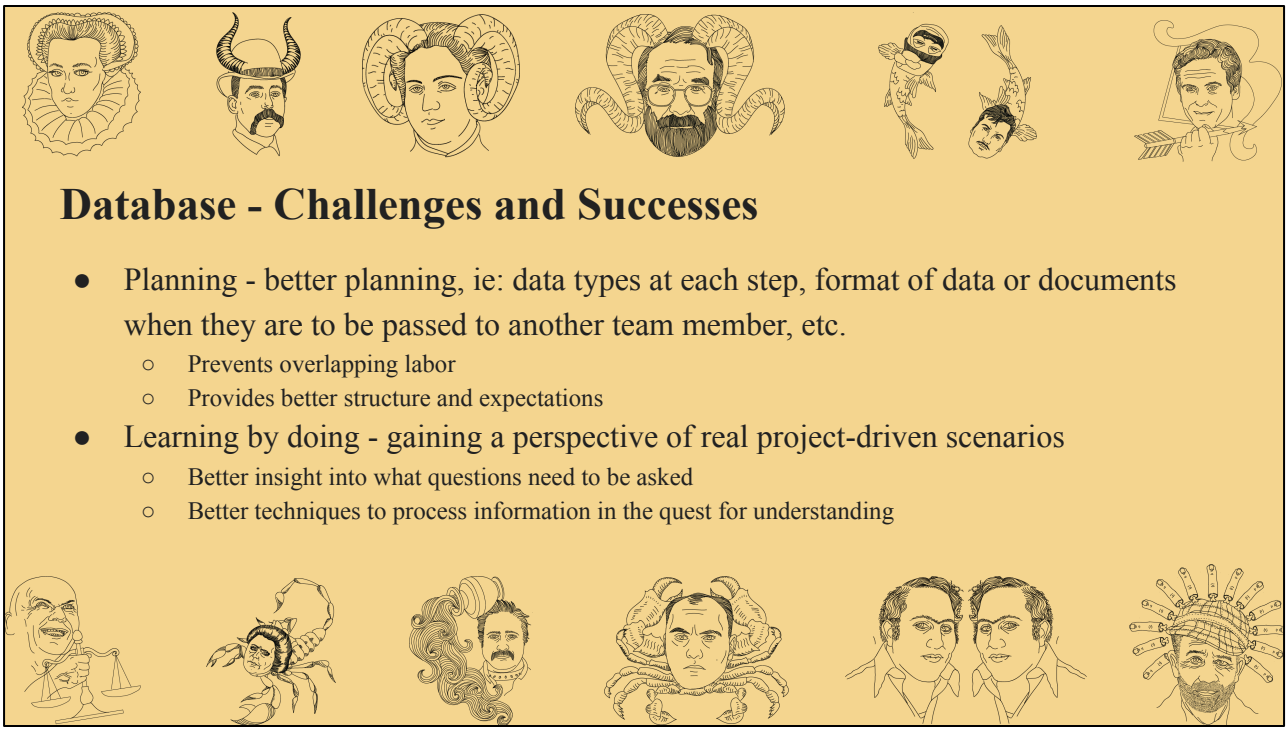


Angelique: Discuss our initial debate between what database type to go with, and why we chose an AWS PostgreSQL database.Once we had found our data, we needed to organize it and store it. For storage we opted for hosting on AWS. The Amazon RDS (Relational Database Service) provides dynamic storage that interfaces directly with PostgreSQL. With a connection to the host, all team members could access the data in PGAdmin through a PostgreSQL database. Although most of the cleaning and organizing was done primarily with Pandas and Python but a small amount was also accomplished in PostgreSQL. It's is not only helpful for the team members to have direct access to the data, it also removes the need to send large data files by email.



Angelique: Feel free to elaborate here if you want to.

Once the database was built and connected to AWS, tables were built and data was loaded as a test. An inner join was used to combine data from separate tables. A connection string was created. Using a connection engine in the the webscraping file and the machine learning file, the database was integrated into the ETL.



Angelique: Elaborate on some of the struggles we had with the Database and describe how we were finally able to perform an inner join after formatting the date columns for the full moon chart and the KCPD Crime data.

Challenges

Planning was the most important thing we could have done differently. I caught myself doing work that didn't need to be done or had been done. There were also some junctures in the work that could have been planned out better.

Ex 1. Cleaning data that got removed.

Ex 2. Cleaning data that is cleaned.

Ex. 3 Clearly defined formats for all work that is to be passed to another team member.

Successes

Learning to have some insights into the workflow. Knowing more in-depth questions to think about in the planning phase of analytics projects.

Analysis Phase - Wikipedia Data

Lauren: We put the data into tableau and excel pivot tables to examine any trends between astrological sign and the number of serial killers. We found there were more aquarius killer than other types.

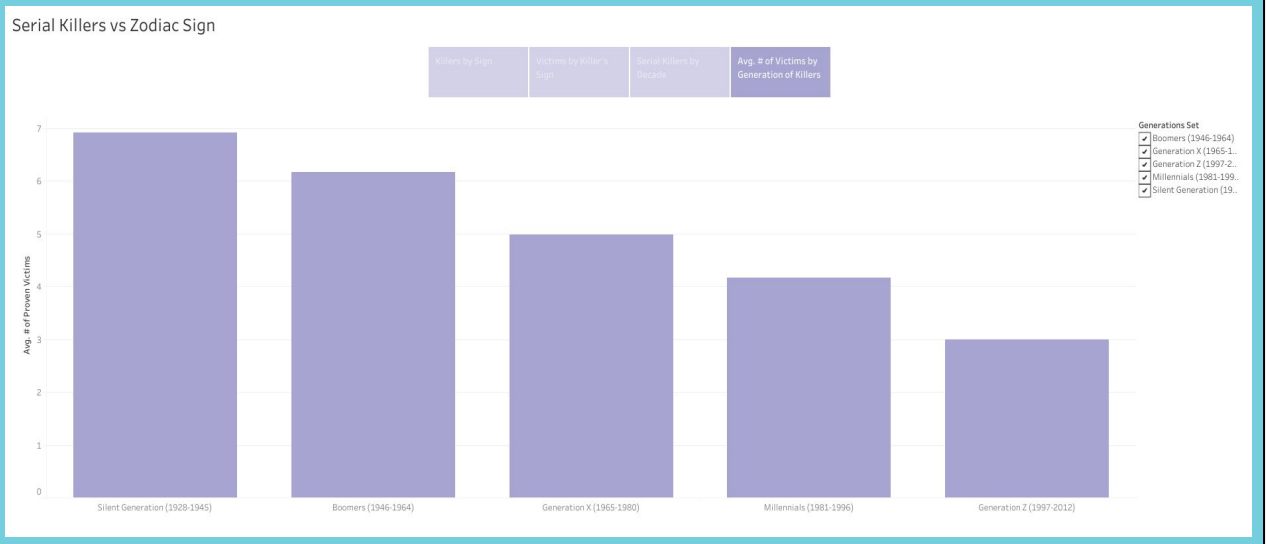
Analysis Phase - Wikipedia Data

Lauren: Then, we analyzed number of victims (confirmed) by Killer's astrological sign. Again, the highest number was from aquarius.

Analysis Phase - Wikipedia Data

Lauren: Next, we separated the killers by their birth year. We found that the highest number of serial killers were born between 1946-1964 (with the most being in 1953). This generation definition comes from the Pew Research Center.

Analysis Phase - Wikipedia Data

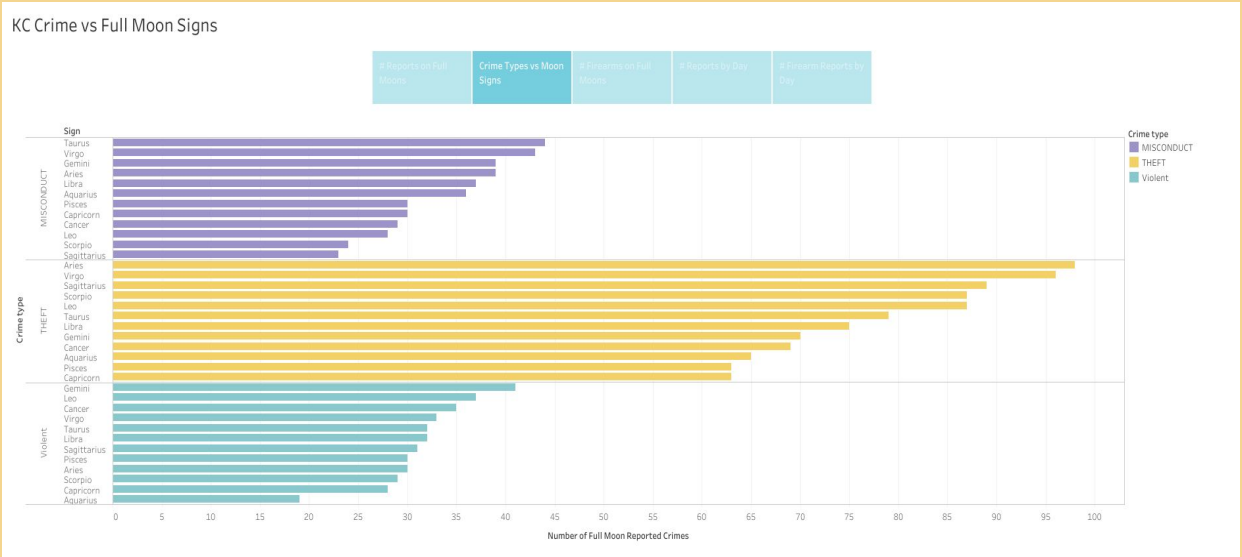


Last, we looked at the average number of victims associated with each generation of killer. The average highest number of victims (6.907) came from the silent generation, which were individuals born between 1928-1945.

Analysis Phase - KC Crime vs. Full Moon Data

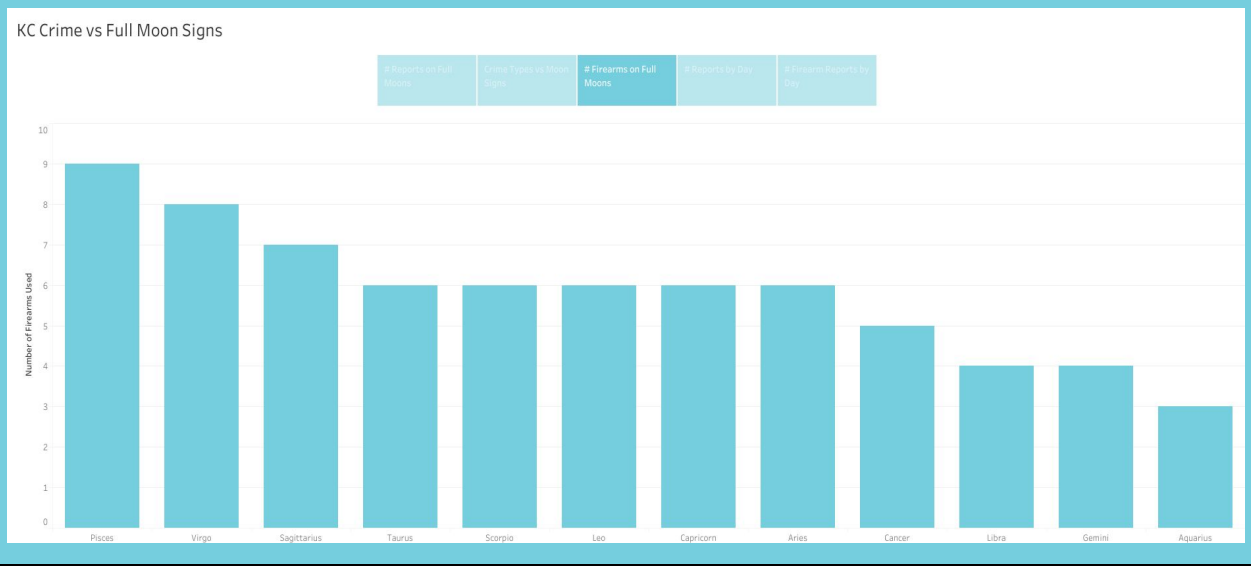
Lauren: First, we looked to see which astrological season full moon was associated with the highest number of crime reports for 2009. Virgo full moon (associated with September) had the highest number of reports.

Analysis Phase - KC Crime vs. Full Moon Data



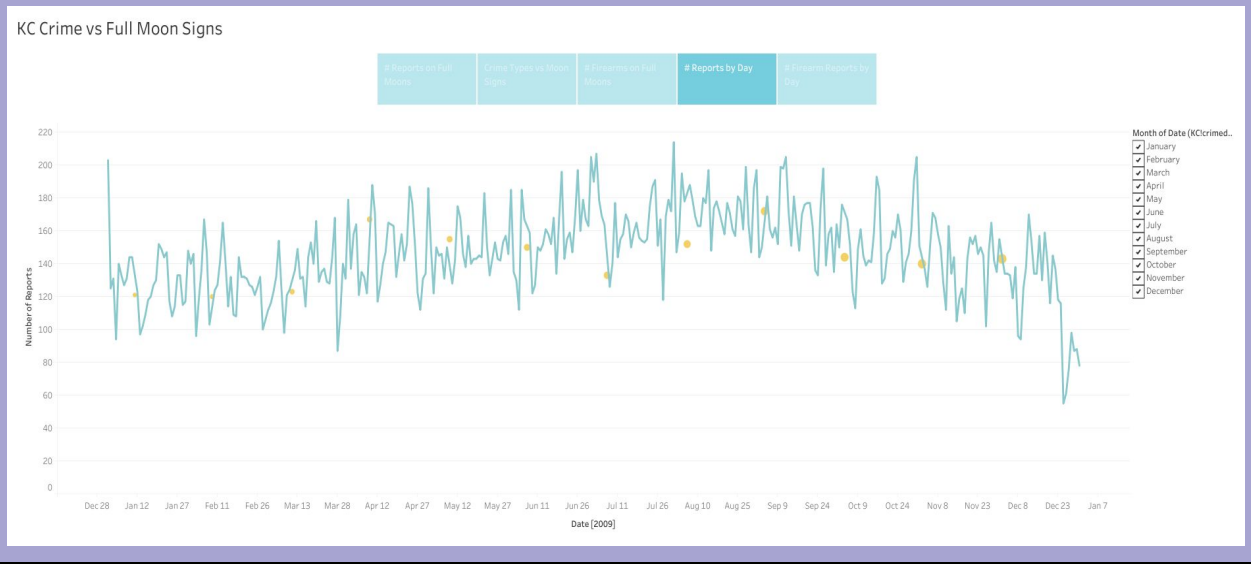
Then, the reports were separated by crime type to see which astrological season had the most for each crime type. Crimes categorized as "misconduct" were most often committed during the Taurus full moon, theft was committed most during the Aries full moon, and violent crimes occur most often during the Gemini moon.

Analysis Phase - KC Crime vs. Full Moon Data



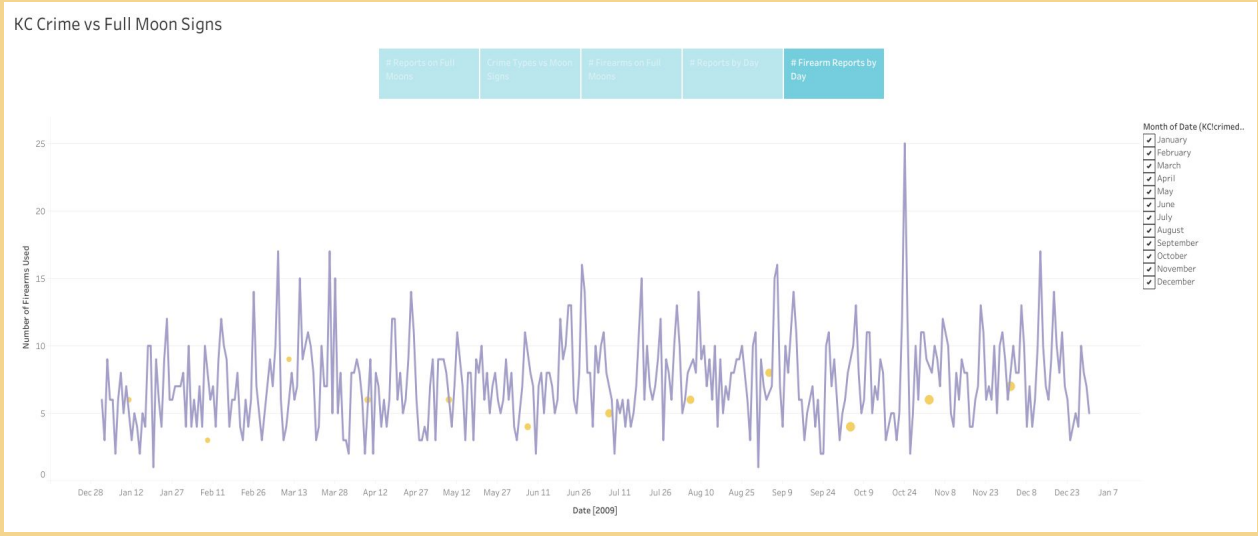
Lauren: Next, we looked to see which full moon sign had the most reports associated with firearms. Pisces full moon had the most firearms.

Analysis Phase - KC Crime vs. Full Moon Data



Lauren: Then to better see if there were spikes during full moons, all of the crimes were placed on a line chart with the X-axis being the date the crime was committed and the Y axis being the number of reports. The graph could be filtered to show each month individually. In general, there did not appear to be a major uptick in reported crimes on full moons.

Analysis Phase - KC Crime vs. Full Moon Data

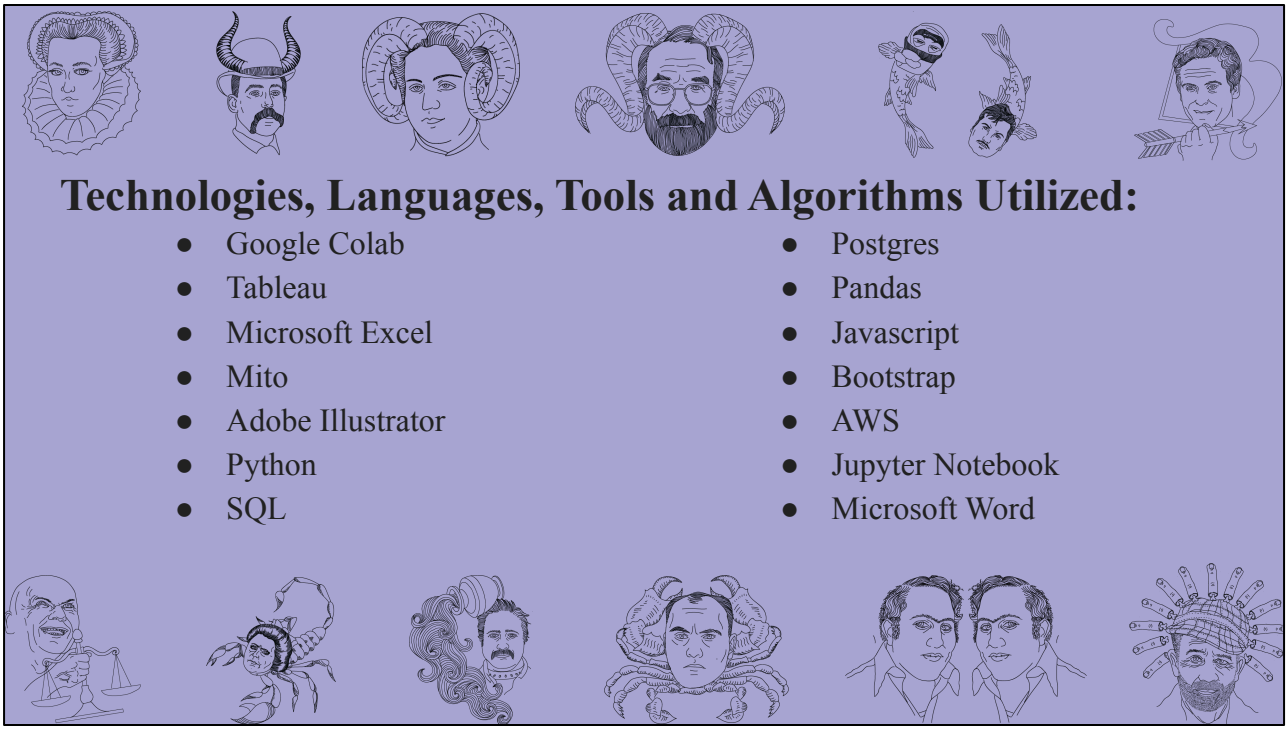


Lauren: Last, the same line chart was created to see if more firearms were used during full moons. Again, in general there did not seem to be a major increase.

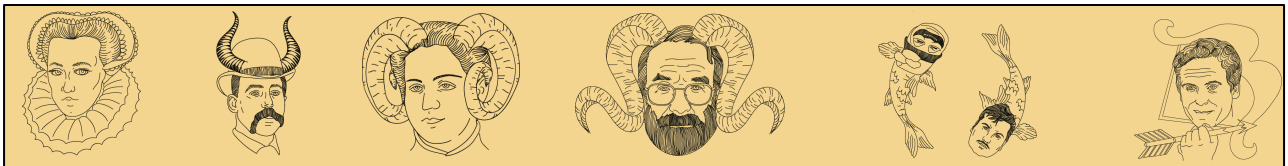
Analysis Phase - Tableau Presentation

Tableau Link Here

Lauren: Either Live Demo the Tableau Charts or proceed with following charts.



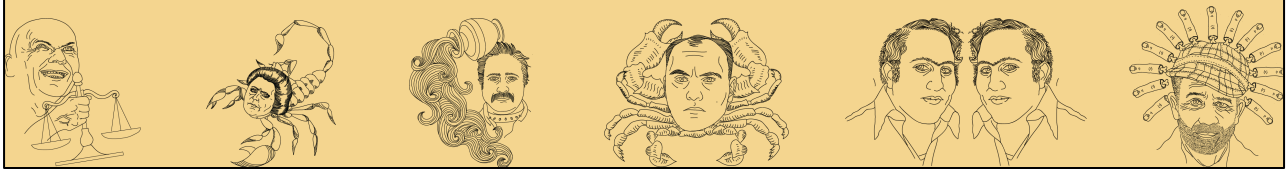
Lauren: Discuss or list technologies utilized.



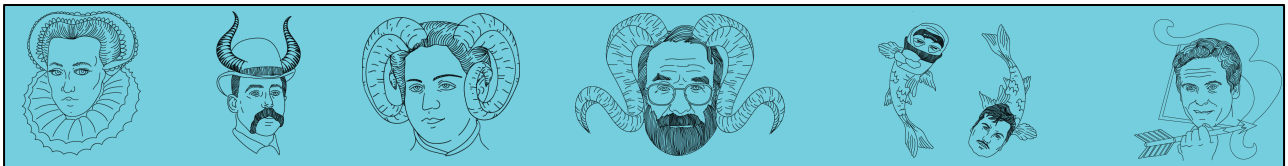
Results of Analysis - Machine Learning Models

- Logistic Regression models were used as both datasets were categorical.

- Unfortunately, both models showed a low accuracy score as well as precision, recall and F1 scores.

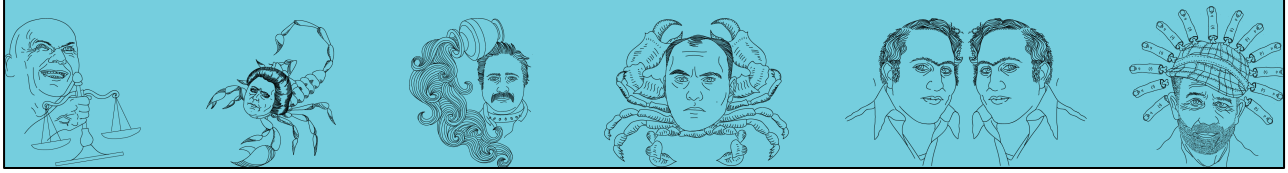


Jessica: (Explain basic setup of our machine learning models, feel free to show examples of code utilized if wanted.) Because both datasets were categorical, we decided to use Logistic Regression Models for Machine Learning. Our hope to prove a correlation between the zodiac signs and crime were sadly not validated using the Logistic Regression models.

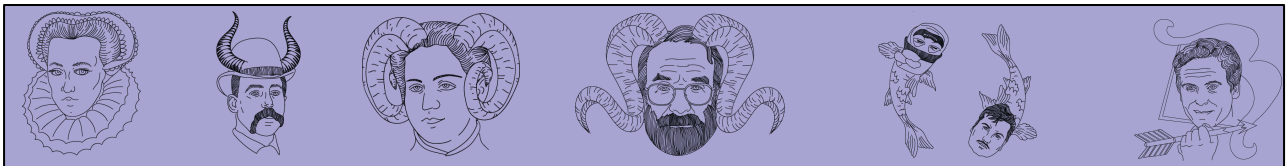


Results of Analysis - Machine Learning Models

- Question: Does Crime happen more often during certain astrological seasons?
- Answer: The results in the Classification Report for the KC Crime Dataset show that our Logistic Regression model CANNOT accurately predict the answers we were looking for on this data set. Crime does not occur more often during a certain astrological season.



Jessica: The accuracy of the model on the Kansas City Crime dataset was only 11% after making some adjustments to the original model's train_test_split and standardizing the data. A shuffle of the data was also performed to test for better results in case the issue was with how the data was organized before it was run through the model.



Results of Analysis - Machine Learning Models

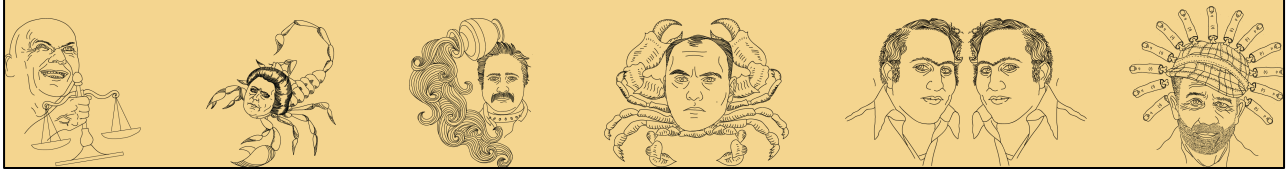
- Question: Is there a particular astrological sign more likely to be a serial killer? Does date of birth relate to number of victims or number of killers?
- Answer: The results of our Classification Report indicate the precision scores are all 14% and below, meaning the Logistic Regression model CANNOT accurately answer our questions.

Jessica: The same attempts for more accuracy were made to the Serial Killers dataset. Unfortunately the results were similar. The accuracy score topped out at 14%.

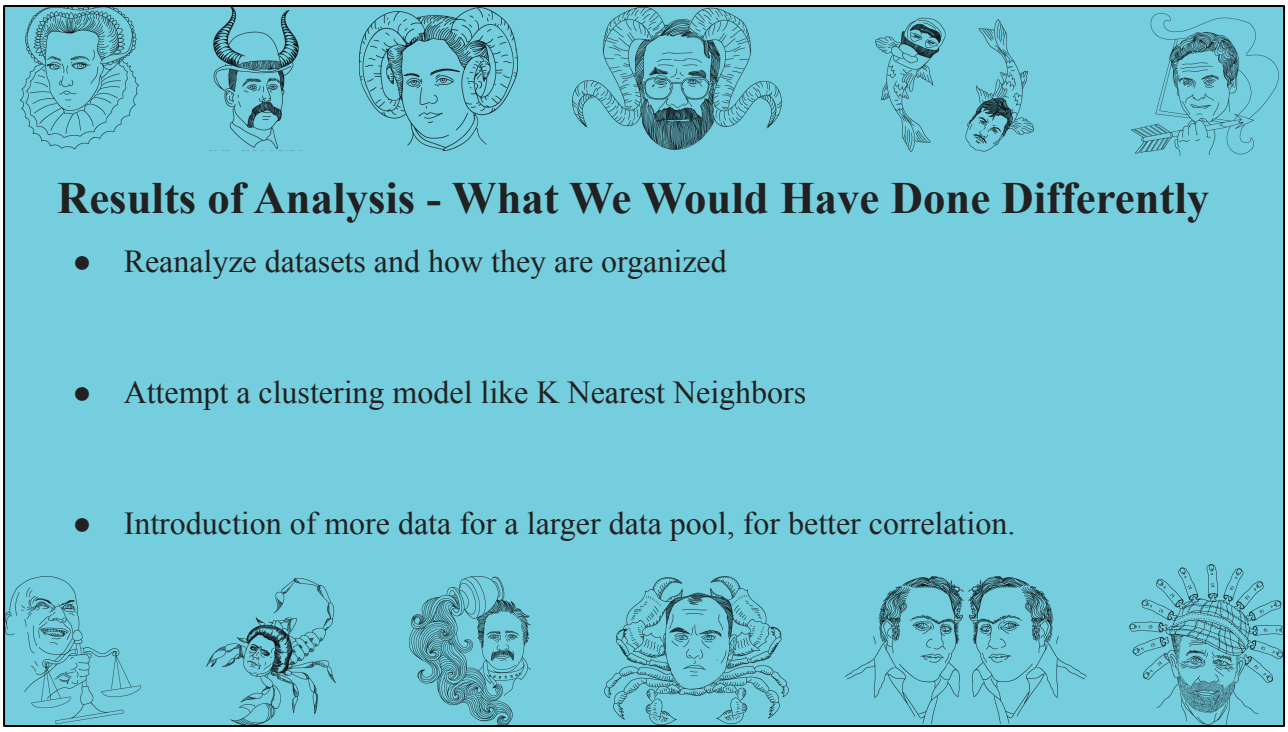
Precision, recall, and F1 scores for both models were also quite low. Likely no model could sense it likely makes logical sense that certain astrological signs are not associated with a higher likelihood of murder.

Results of Analysis - Machine Learning Models

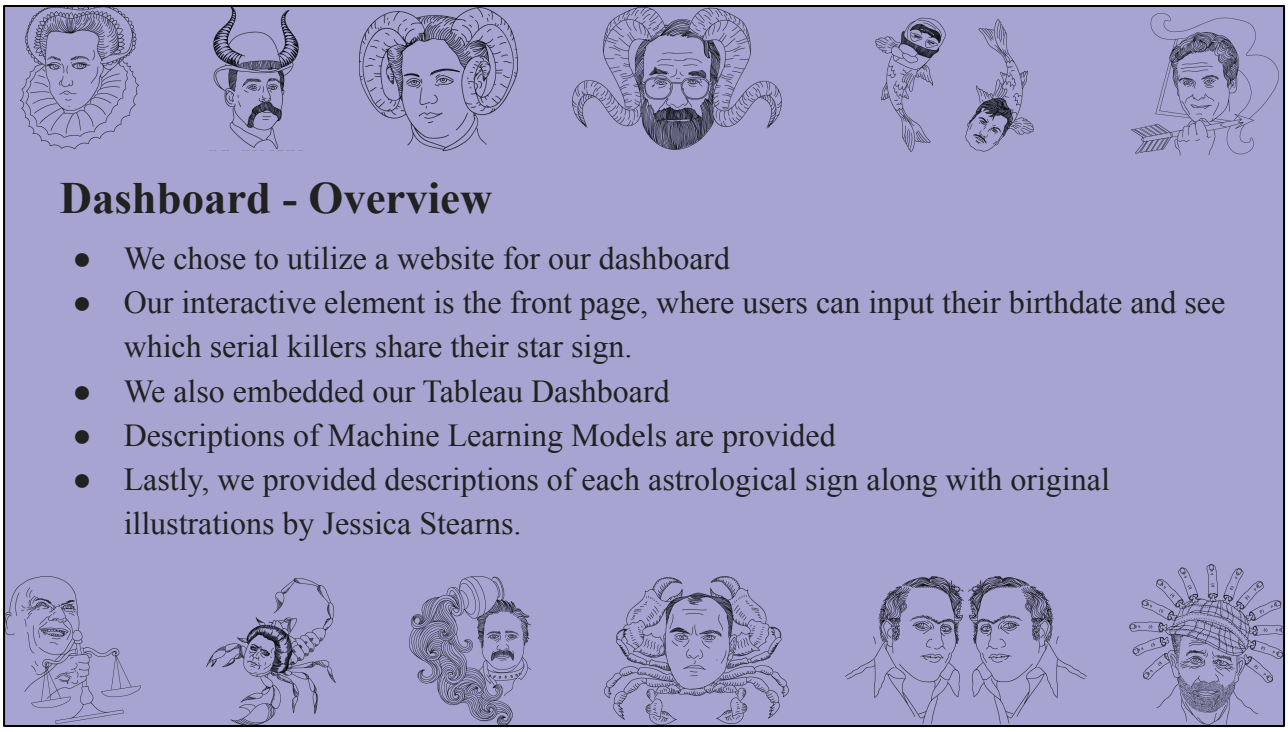
- Question: Are Pisces men the worst?
- Answer: Despite little-to-no support from our Machine Learning Models, we stand by our hypothesis that Pisces men, are in fact, the worst.



Jessica: Feel free to ham this one up or elaborate further, or just skim it lol.



Jessica: If this analysis were to continue, the datasets should be re-analyzed to make sure they are not set up in any sort of order that can skew the machine learning models. Attempting a cluster model like K Nearest Neighbors may show more accurate results on both datasets. Changing the features or introducing new data to the datasets to use as new features could also change the outcome of the Machine Learning performed during this project. For crime it would also be better to include more data, a wider range of cities, and over more years.

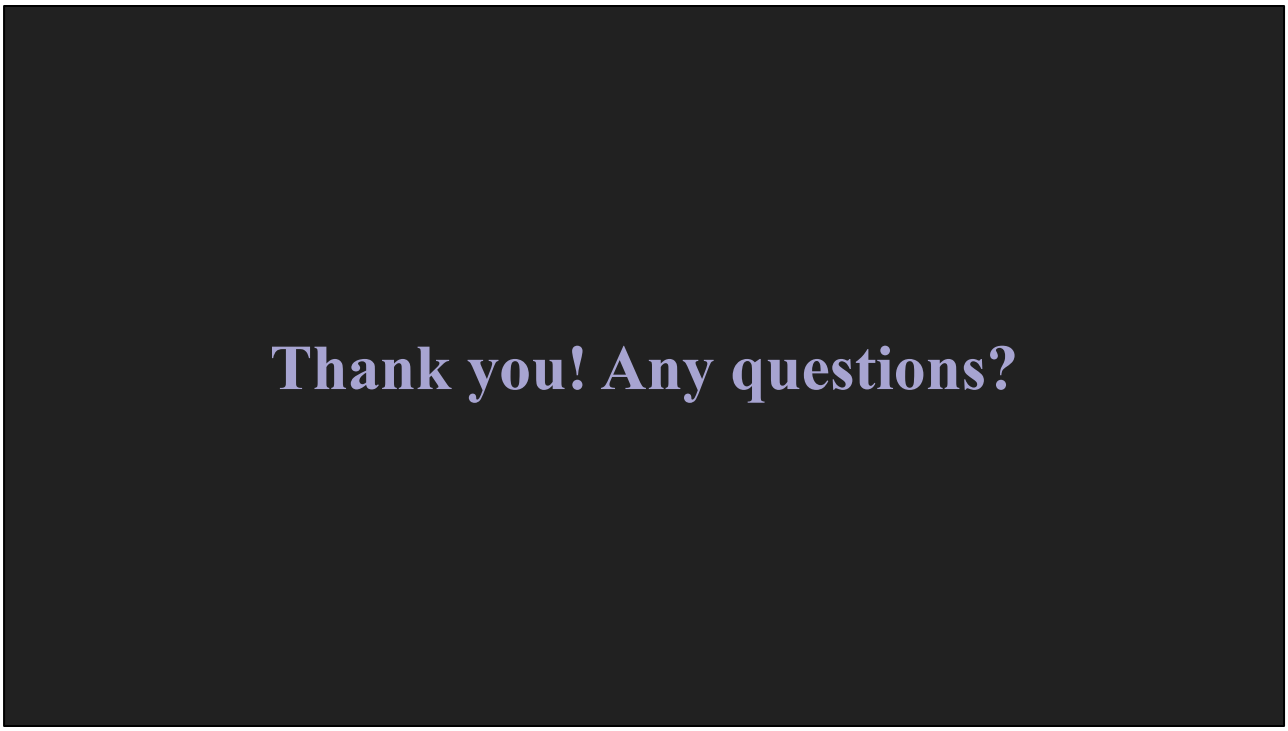


Dashboard - Overview

- We chose to utilize a website for our dashboard
- Our interactive element is the front page, where users can input their birthdate and see which serial killers share their star sign.
- We also embedded our Tableau Dashboard
- Descriptions of Machine Learning Models are provided
- Lastly, we provided descriptions of each astrological sign along with original illustrations by Jessica Stearns.

Corrine: Just briefly graze these notes and elaborate where you would like to.

Corrine: Please perform a live demo of the website at this slide.
https://rulamia.github.io/Horrorscopes/



Anyone: Thanks to all fellow students and instructors for a great 6 months!