Lindsey **Maya** **Meredith** **Rylee**

Group 1: Avalanche_Analysis
10 min to present

Slide: 1

Greeting:

- Hi, we are group one! - Names

Slide: 2

Opening:

- The framework for this project was to analyze avalanche data from Snowbound Solutions LLC based out of Boise, ID and present our findings to the owner, Scott.

Slide: 3

Question:

- What weather features contribute most to Avalanche occurrences in Juneau, Alaska?

Slide: 4

Why We Selected Topic:

- (purpose slide has info)

Slide: 5

Resources:

- This data was presented to us from Snowbound Solutions LLC
- It included 6 files. We ended up focusing on 2 of the documents for our exploration: The Daily_obs and the avalanche_obs.
    - The daily_obs doc included daily weather observations recorded by an expert from various locations in Juneau, Alaska
    - The avalanche_obs doc included data relating to recorded avalanches in the area from 2019-2021

Slide: 5-6

Database Building:

- Rylee and I took on the task of organizing and weeding through our data
- We used a mix of Postgres and pgAdmin and python/ pandas for this process

Lindsey **Maya** **Meredith** **Rylee**

- Rylee first converted our original data which was mostly in json format to csv and dropped some unnecessary columns and null values to make the initial table creation in pgAdmin easier.
- I then used pgAdmin to build tables of our datasets. I built tables for the daily_obs and avalanche_obs datasets.
- Finally, I did a outer join on those two tables based on date to create a dataset that we would use in the machine learning model

Slide: 7

Data Preprocessing:

- After we had our data, we created a connection string using SQLAlchemy to connect our database in Postgres to our notebook for use in our model.
- For the preprocessing stage to make sure our data was ready for our model, we:
    - Converted the avalanche_occurred column to a yes/no binary column by replacing null values with no as well as replacing the observation dates with yes.
    - Then, dropped null values in the data frame
    - Next, we dropped the observation dates column
    - After that, we encoded our categorical columns such as the wind direction, sky cover, precipitation type, etc.
    - Lastly, we scaled the data which is super important when training the model and giving each feature the same footing without any upfront importance.

Slide: 8-10

Machine Learning Model:

- Initially we decided to explore a logistic regression supervised model in the hopes of predicting whether an avalanche would occur or not based on daily weather observations. This was promising until we ran into issues merging the daily observations table with the avalanche occurrence table.
- We then decided to take an exploratory approach and worked on a k-means unsupervised model using PCAs. We hoped that instead of predicting avalanches, we could instead determine what conclusions and patterns an unsupervised model would draw from our data. We got the

model to work but decided that the results did not tell us anything new or help us in answering our main question.

- Finally we landed on a feature importance supervised model. This model tells us which daily observation features contribute most to the occurrence of avalanches, which felt very fitting for the kind of exploration we set out to do.
- Here are the results of the feature importance model, the score represents how significant each feature is to avalanche occurrence. Here you can see that air temp min, max and current, as well as hazard score are the most weighted features. Seeing which features were most important allowed us to know which features to further analyze and visualize.

Results and Visualizations:

Slide: 11-16

- FLIP TO TABLEAU STORY
- Avalanche Occurrence by Quarters: This graph is a breakdown of avalanche counts by quarters from 2019-2021. Most avalanches occurred during Q1 with an exception of 2020, a heavy avalanche year in other quarters.
- Air Temperatures by Quarters: The graph supports the avalanche occurrence by quarters by having an average lower temperature in Q1 which is the winter months of January-March.
- New Snow Height by Quarters: The new snow height graph aligns with the air temperature and avalanche occurrence with the most snow height accumulated in Q1 and resulted in the most avalanches occurring.
- Counts of Different Precipitation Types: When looking at the different types of precipitation, 179 avalanches occurred when there was no precipitation and 49 avalanches happened when there was snow.
- Counts of Different Sky Cover Types: This is the breakdown of different sky cover types you can see that avalanches occurred most on clear and overcast sky days.
- Hazard Count: This graph shows the hazard level on days where avalanches occurred, surprisingly the most avalanches reported were on a hazard level 2.

Lindsey **Maya** **Meredith** **Rylee**

Slide: 17

Summary of what we learned:

- In summary, we know that none of this is revolutionary but it is good to have this analysis as a backup to answer the question we were originally asking. We found that the features that weighed most heavily were air temperature, precipitation, sky cover and new snow height. When analyzing the different types of precipitation and sky cover type it was found that the most avalanches occurred when there was no precipitation and clear skies, which makes sense because this typically means the sun is out making the snow unstable. The most avalanches occurred in quarter one of the three years recorded which didn't come as a surprise since this is the winter months. We were amazed that even when a trained professional was predicting hazard levels, avalanches were still occurring when rated a low hazard. This goes to show that avalanches are still highly unpredictable. In conclusion, avalanches can be very dangerous and detrimental to communities when they occur so better understanding what can contribute to these happening can hopefully better prepare these communities in Alaska.

Slide: 18

Further Analysis:

- If we had the time we would have liked to map this data to show the different geographical areas covered in Alaska and create a more advanced learning model that shows a more long term accumulation that can cause avalanches.

Closing:

5 min for questions

Slide: 19

- Thank you so much for sitting through our presentation.

- Does anyone have any questions?
    - FLIP TO TABLEAU STORY