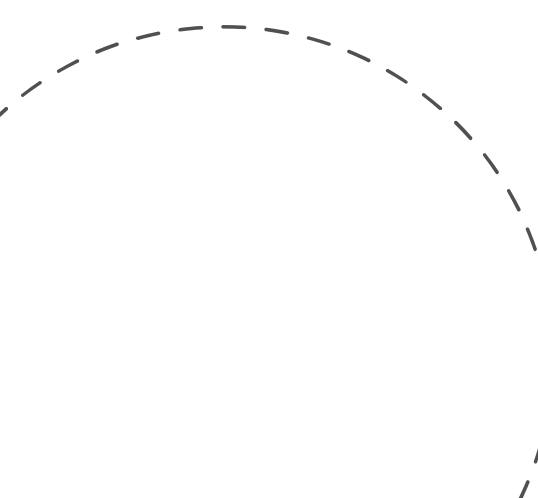


ETHEREUM PRICE & SENTIMENT

Bausch Koh, Cynthia Zhang, Elaine Cui



Our Team



Bausch Koh
Class of 2024



Cynthia Zhang
Class of 2025



Elaine Cui
Class of 2024



Agenda

Purpose

Motivations in selecting our project idea

Approach

Our process in tackling the problem at hand

Data Exploration

Dissecting key features of the dataset

Analysis

Utilizing analytical tools to discover meaningful trends

Insights

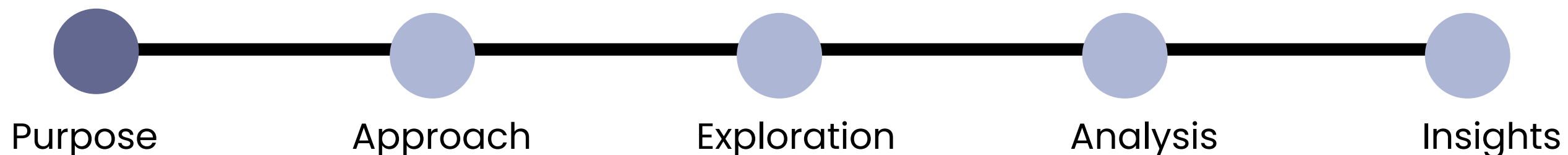
Interpretation of results and future implications



Purpose

As cryptocurrencies gain traction in **real-world applications** such as decentralized finance, investment products, and payment methods, our team wanted to examine the possible factors behind fluctuations in price of a cryptocurrency coin.

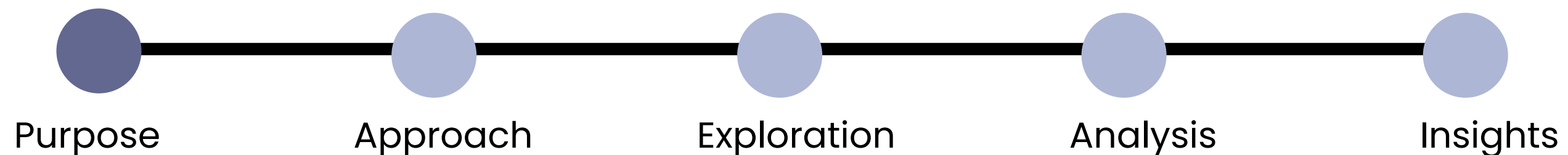
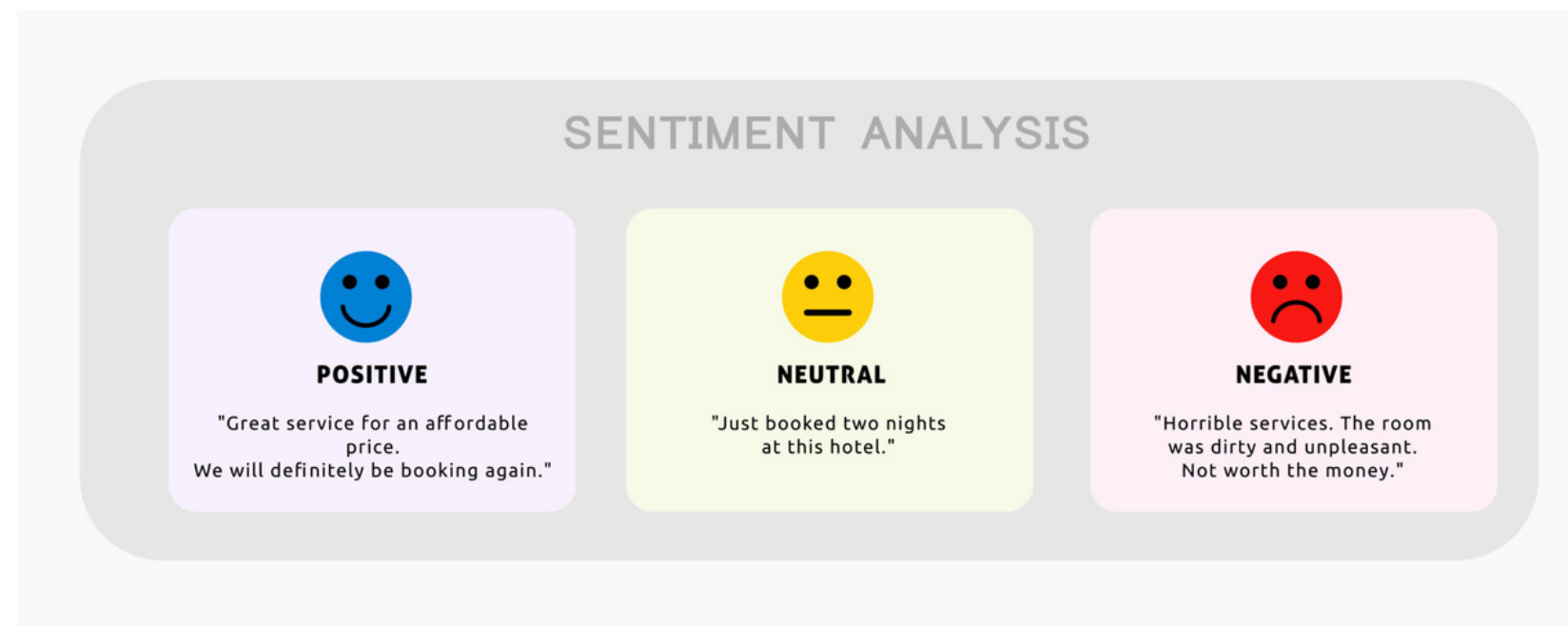
ETH is the second largest cryptocurrency by market cap at \$342B, and is increasingly important in online transactions, including NFTs sales. Our team decided to utilize data on ETH price and sentiment score analysis to **identify potential factors behind price fluctuations.**



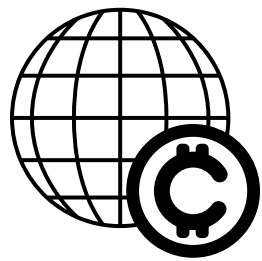
Background

What is Sentiment Analysis?

Sentiment analysis, also known as “opinion mining” or “emotion artificial intelligence”, is an approach to natural language processing that identifies the **emotional tone** behind a body of text.

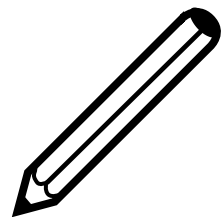


Data Summary



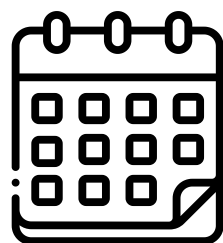
Background

This dataset provides historical data of the top cryptocurrency coins



Creator

Dataset created by Sundal Kumar on Kaggle



Time Range

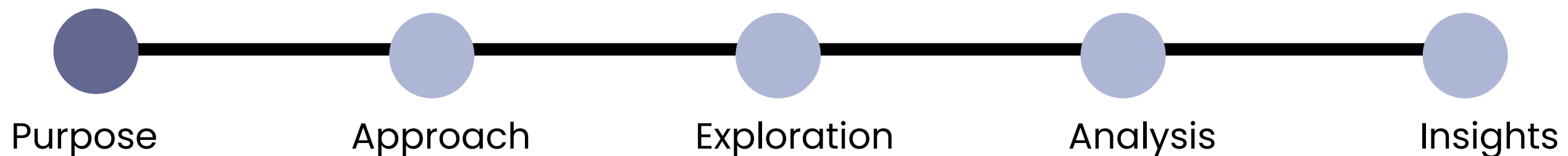
Data provided on a daily level, with date range from 2015-2021

of Rows

2160

of Columns

10



Approach



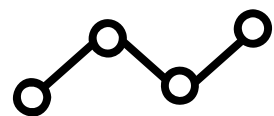
Reddit

We scraped relevant **Reddit posts** over the time period covered by our ETH data for their title and date posted with a **web scraper** to examine global attitudes toward Ethereum



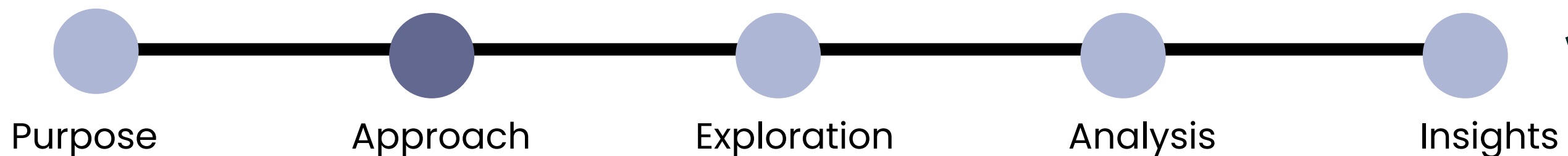
Sentiment

We used IBM Watson for sentiment analysis to analyze the titles of Reddit posts for their **sentiment score** (-1 being the most negative, +1 being the most positive)



Regression

We conducted **linear and logistic regression** analyses on our ETH and sentiment data to understand how price fluctuations correlate with sentiment



Sample Reddit Posts

↑ Posted by u/[deleted] 7 years ago

22 **Why should the average person care about Ethereum?**

↓

As I see it sounds good to technologists but I see no reason why the average person would want to use Ethereum.

Can anyone explain what benefits it would provide?

↑ Posted by u/coinop-logan 5 years ago

152 **I just burned 1 ETH on purpose! For my next trick, I've committed to losing another 15 ETH, to demonstrate how burn-capable smart contracts can incentivize arbitrary service requests to be filled by reliable strangers.**

↓

↑ Posted by u/[deleted] 4 years ago

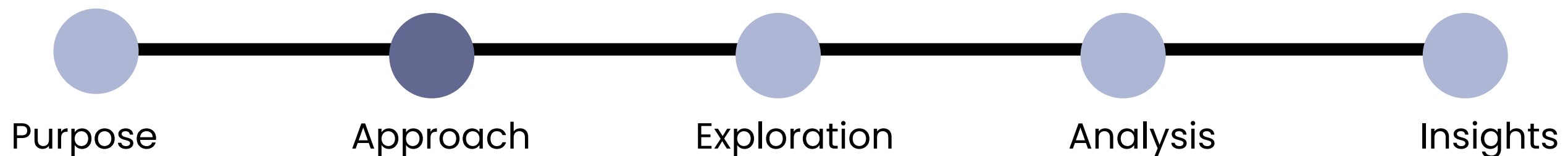
123 **I bought McDonald's using Eth!**

↓

Sort of...

↑ Posted by u/Mortum1 3 years ago

356 **The more I read about ETH, the more I believe it's far superior to BTC.**



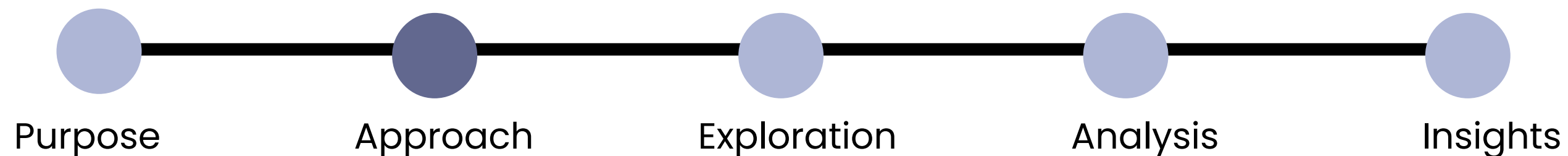
Web Scraping Process

BeautifulSoup

Beautiful Soup is a Python library for parsing structured data. It allows you to interact with HTML in a similar way to how you interact with a web page using developer tools.

For web scraping to work in Python, we performed three basic steps:

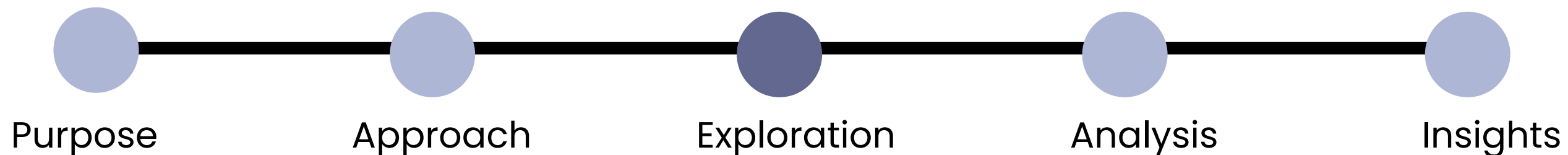
1. Extract the HTML content using the requests library.
2. Analyze the HTML structure and identify the tags which have our content.
3. Extract the tags using BeautifulSoup and put the data in a Python list.



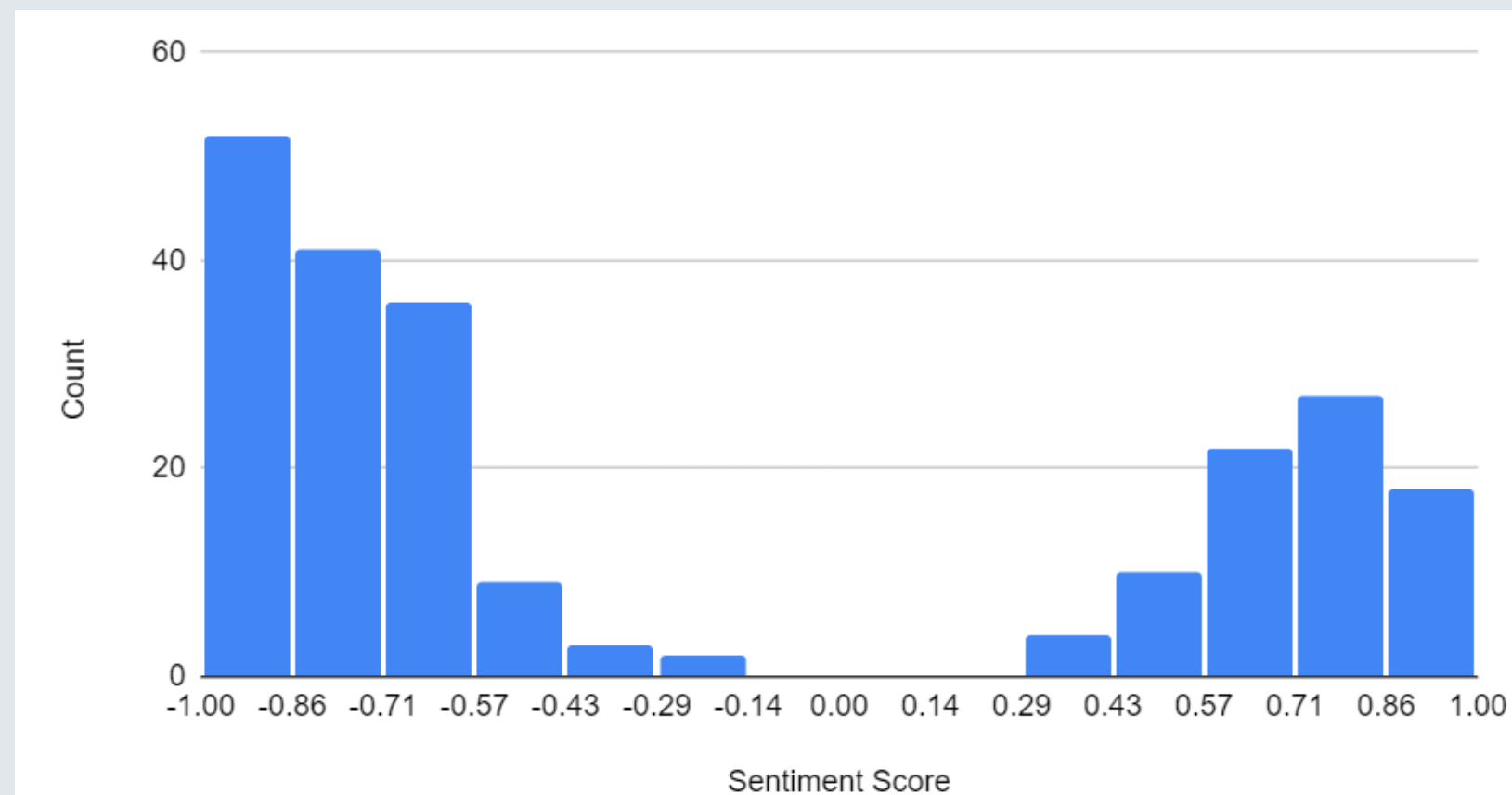
Sentiment Results

We excluded results with a sentiment score of 0 because they indicated posts without strong emotion, which would not be useful for our analysis.

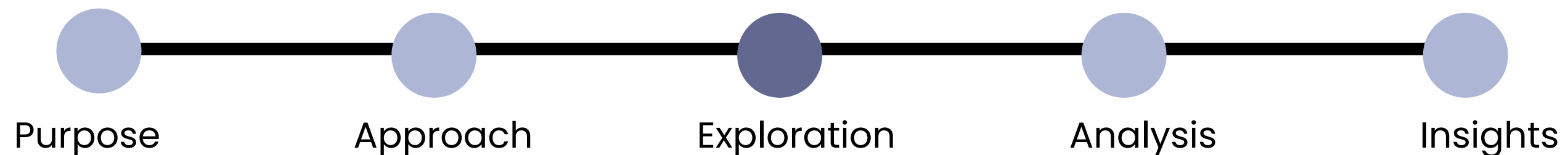
Title	Link	Sentiment	Date
ETH Mining is highly profitable, and network is currently only secured	https://www.redd	0.737746	08/11/2015
Do I need to re-lock my account after sending a transaction? How?	https://www.redd	-0.746093	08/11/2015
Ethereum birthday cake	https://www.redd	0.695804	08/12/2015
Why does My address showing "error Account not found" in ethercha	https://www.redd	-0.884637	08/13/2015
BOUNTY: Open-source Ethereum mining pool.	https://www.redd	0.510945	08/17/2015
Everyone stop using ethpool.org and try a different pool! The operato	https://www.redd	-0.949724	08/19/2015
Payouts every hour, faster pool mining, and nightly 1ETH giveaways!	https://www.redd	0.821706	08/25/2015
Holy shit my eth accounts been hacked	https://www.redd	-0.9972	08/28/2015
ETH is dumping HARD, why? What's the bad news?	https://www.redd	-0.950252	09/11/2015
Ethereum has a P2P Mining Pool now!	https://www.redd	0.800198	09/14/2015
Absolutely excellent interview with Dr. Aeron Buchanan of Ethereum	https://www.redd	0.992994	09/21/2015
It's 2025, Ethereum became a huge success, how will my daily life lo	https://www.redd	0.746489	10/16/2015
Post your wallet Identicon and I'll give 1 eth to the most interesting! B	https://www.redd	0.776209	10/22/2015
Economist: The great chain of being sure about things	https://www.redd	0.90828	10/29/2015
Ethereum Wallet stuck at looking for peers	https://www.redd	-0.891555	11/11/2015
Serious question - Is it true that Ethereum transactions reveal the ide	https://www.redd	0.441704	01/23/2016



Sentiment Results

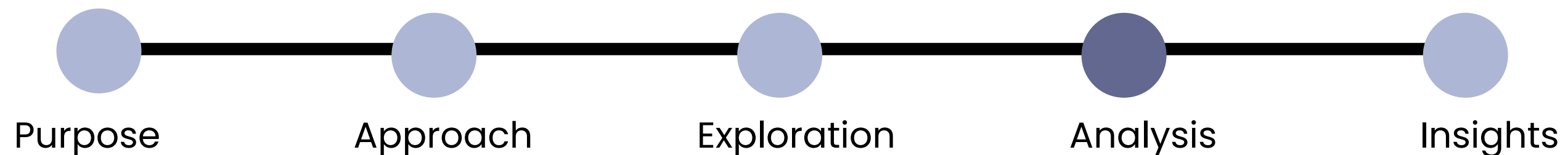
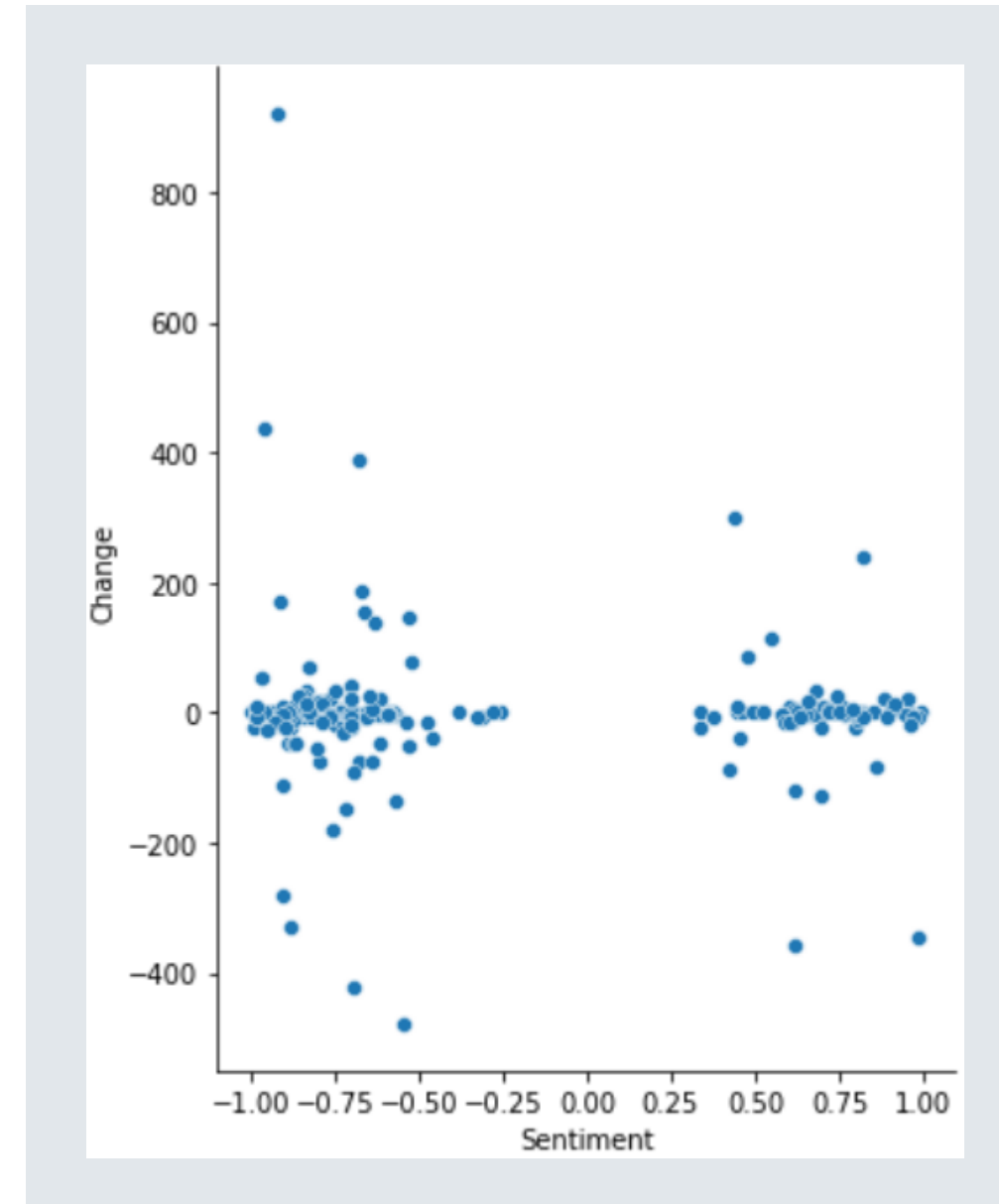


In this histogram visualization, we observe that the Ethereum sentiment scores have a bimodal distribution with a range between -1.00 and 1.00, at $n = 224$



Linear Regression

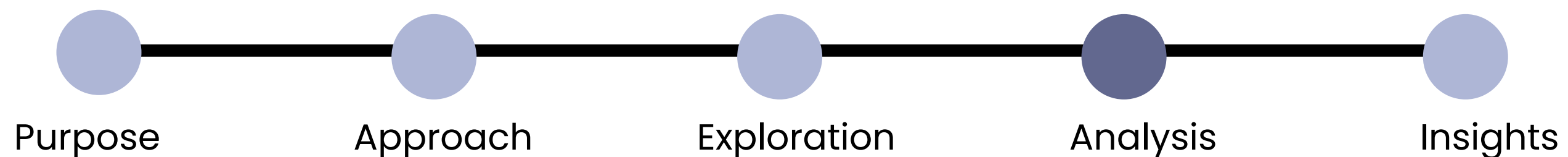
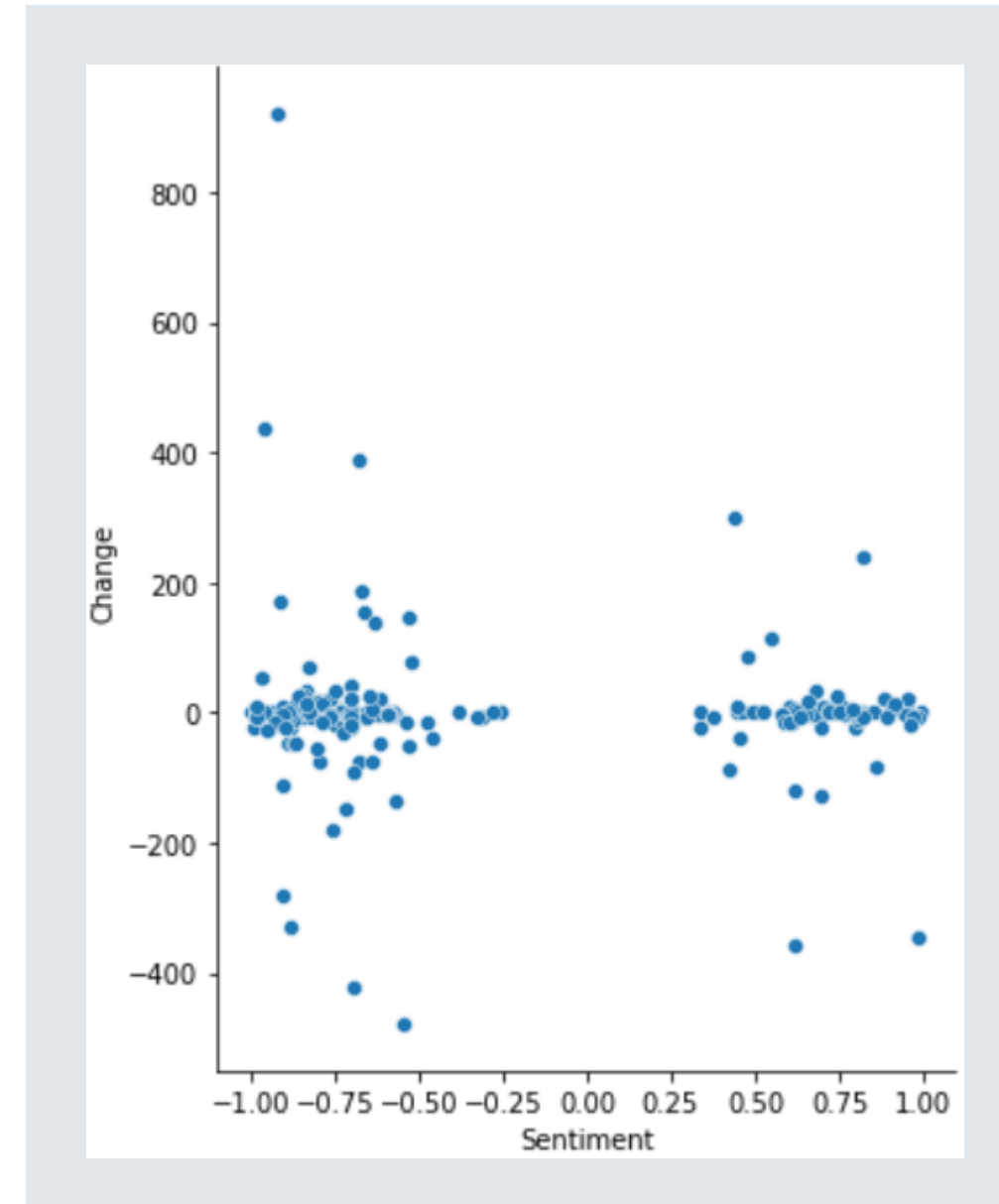
We plotted the sentiment score against the **change in price** for the day, which we calculated by subtracting the closing price from the opening price, then ran a linear regression on the resulting scatterplot.



Linear Regression

Predicted Change in Price =
 $-3.10 - 5.72 * \text{Sentiment Score}$

$R^2 = 0.155\%$
MSE = 11321



Linear Regression Analysis

R-squared

$R^2 = 0.155\%$

Less than 1% of the variation in price is explained by the variation in sentiment

Outlier

There is an outlier in the scatterplot where the change in price was **exceptionally high**, which may have resulted in the negative coefficient

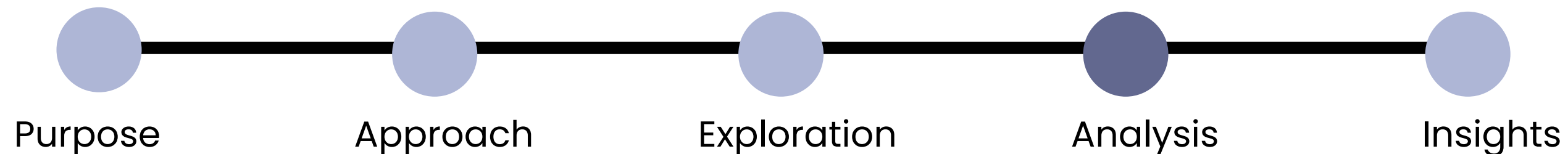
Mean Squared Error

$MSE = 11321$

The large MSE is cause for concern, since **large errors** are not being accounted for by our model

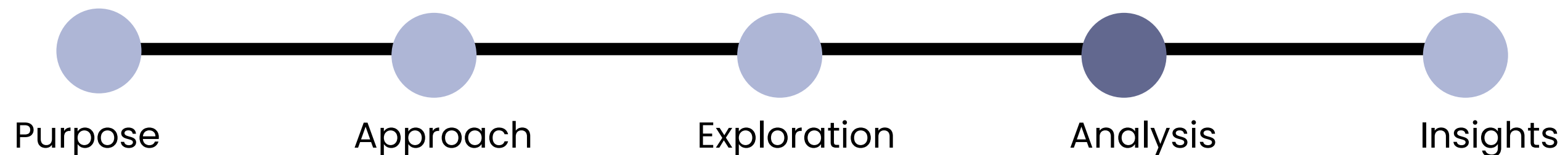
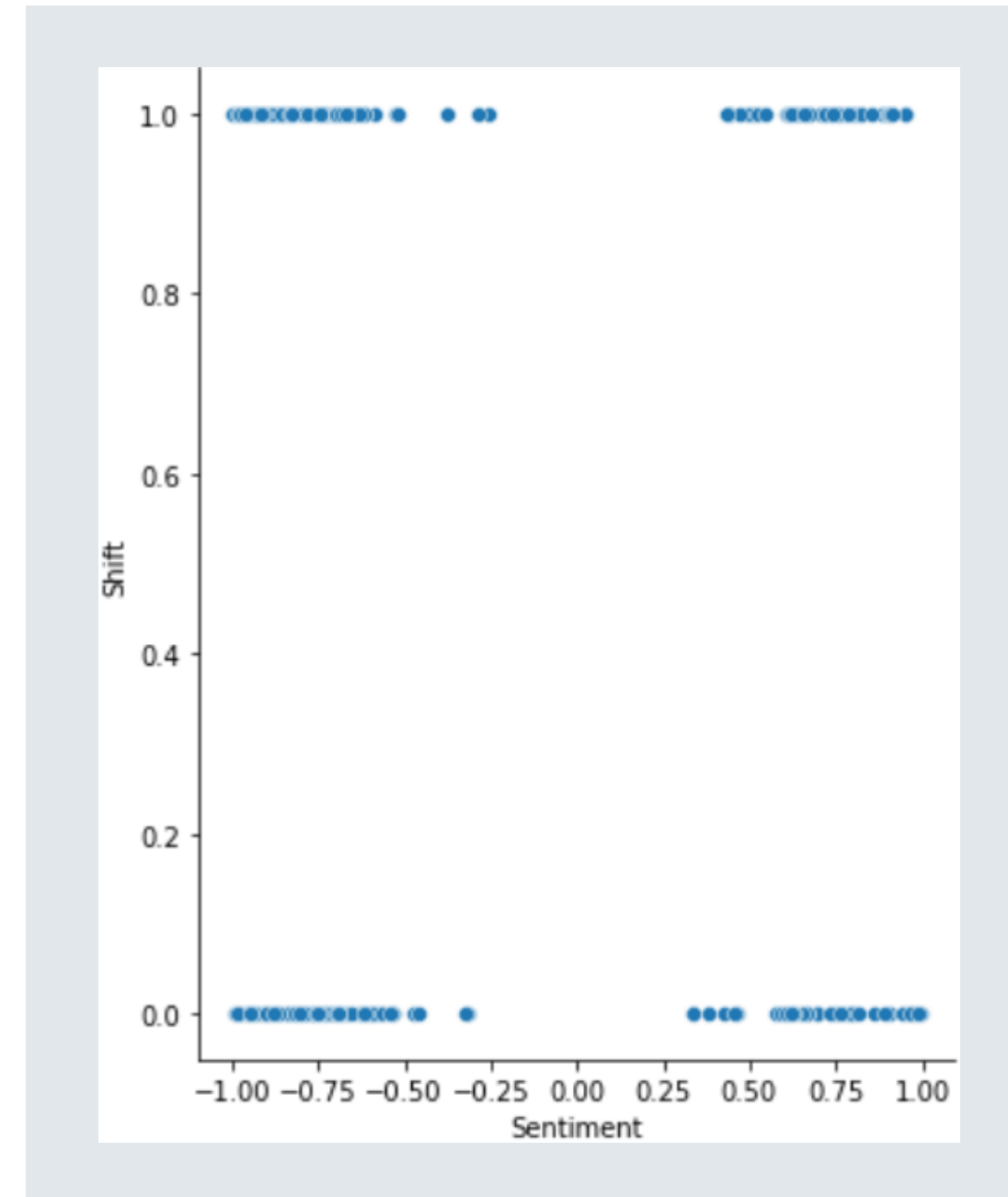
Conclusion

Due to the low R^2 value of 0.155%, we **cannot conclude** a definitive relationship between sentiment and price fluctuations

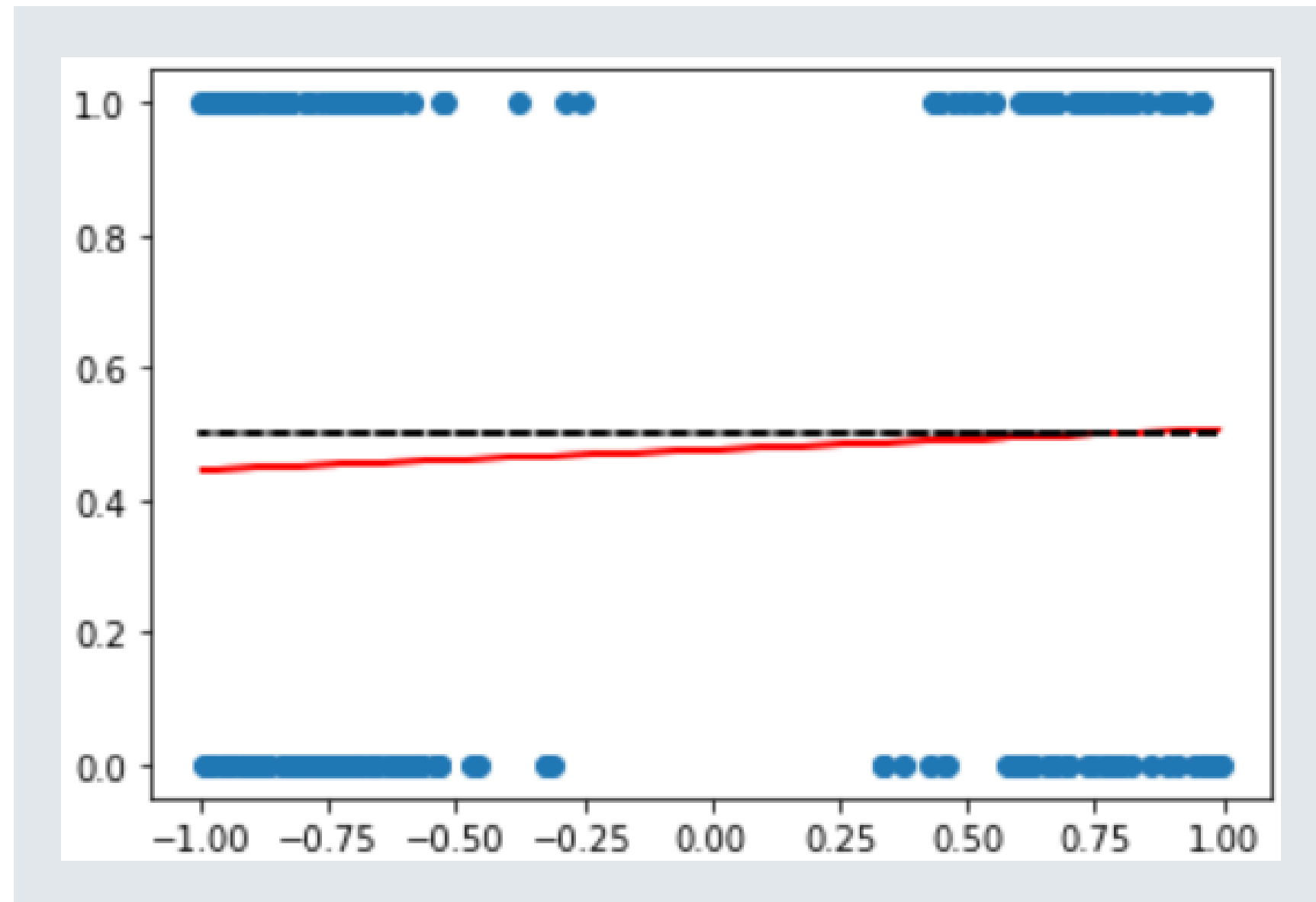


Logistic Regression

We converted the change in price to a boolean: True if the shift was **positive**, False if the shift was **negative**. This allowed us to conduct logistic regression.

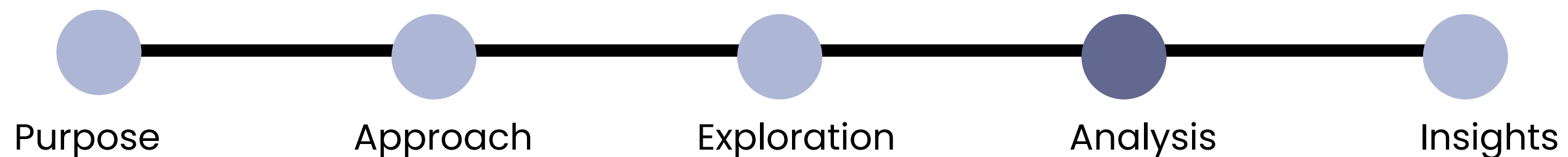


Logistic Regression



Confusion Matrix

[22,] TRUE POSITIVE	[11] FALSE NEGATIVE
[23,] FALSE POSITIVE	[12] TRUE NEGATIVE



Logistic Regression Analysis

Probability

The probability of a positive shift in price for a sentiment score of 0 is 0.475, which indicates that **negative shifts** are slightly more likely

Confusion Matrix

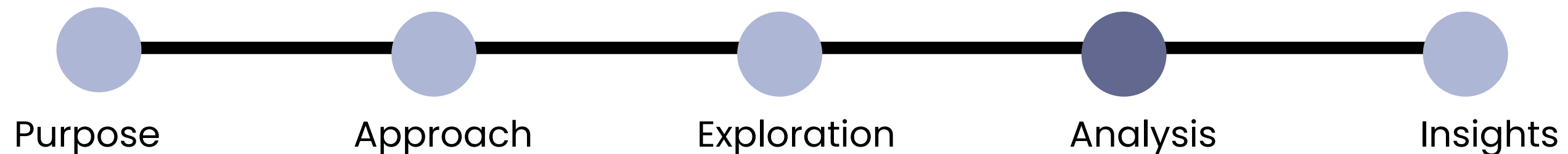
The model predicted around the **same number** of true and false positives as well as true and false negatives

Accuracy

The test set accuracy of the model is 0.5. Since there are only two results (positive or negative), the accuracy is no better than **random**

Conclusion

Due to the **inconclusive results**, we cannot conclude a definitive relationship between sentiment and direction of price fluctuation



Limitations



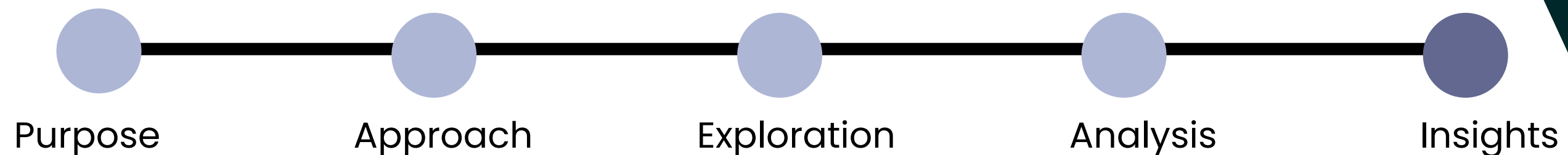
Completeness of Reddit Data

We could not gather Reddit data for every date that was in our ETH dataset. Although we used Google search results filtered by date, to collect sentiment data that was representative of our ETH data's timeframe, we could not ensure that there was a Reddit post created on every consecutive date.



Neutral Sentiment Data

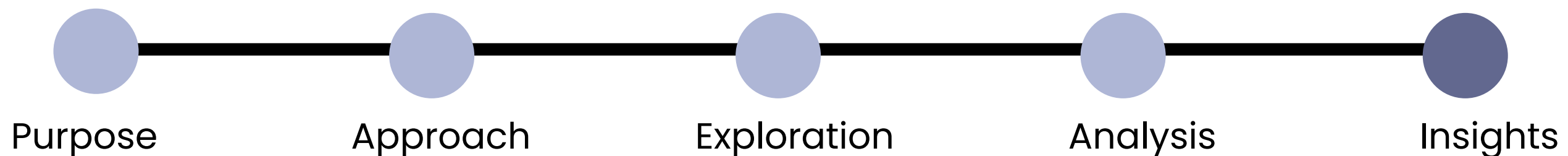
We excluded from our analyses dates with no sentiment data available or a sentiment score of zero. While we made the decision in order to analyze the impact of strong sentiment, we realize that this may have affected the fit of our model, since neutral sentiment data may reflect dates with minimal fluctuation in price.



Conclusion

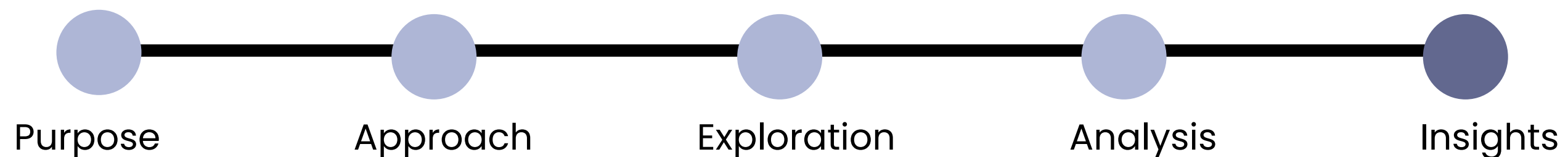
Our linear and logistic regression models showed that user sentiment from Reddit post titles is not correlated with the direction of Ethereum price fluctuations. This may be because prices are so **volatile** that sentiment is not an accurate predictor.

Another possibility is that sentiment is actually a good predictor, but our data was incomplete. This could have happened due to the limitations on the previous slide, or because sentiment data from **another source** (e.g. Twitter) would have been more relevant.



Looking Ahead

The scraping of sentiment scores from social media sites can continue to be useful for investors, who may be looking to **enter the market** at a certain time to take advantage of high or low sentiment. As such, this model provides a framework to help predict possible price fluctuations.





Thank You.
Questions?