

# AI Adoption Analysis

**Trolley Logistics LLC**  
(Ryan Shihabi and Tiffany Le)

## Questions

Can technology commodity prices be correlated to the recent mentions of AI?

---

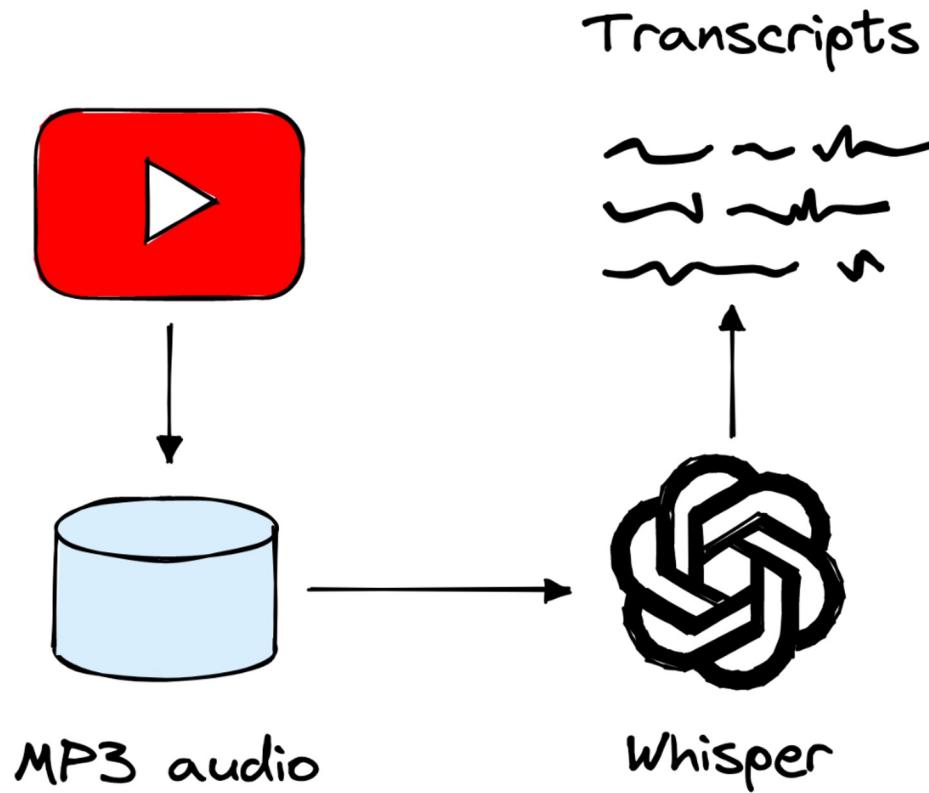
What types of ML tools are being developed?

Can technology commodity prices  
be correlated to the recent mentions  
of AI?

# Data

- Company Keynote Transcripts
  - Nvidia GTC
  - Google I/O
  - Apple WWDC
- Computer and Electronic Product Manufacturing Industry Index (PCU)
- Application Software Publishing Commodity Index (WPU)

# Transcript Collection



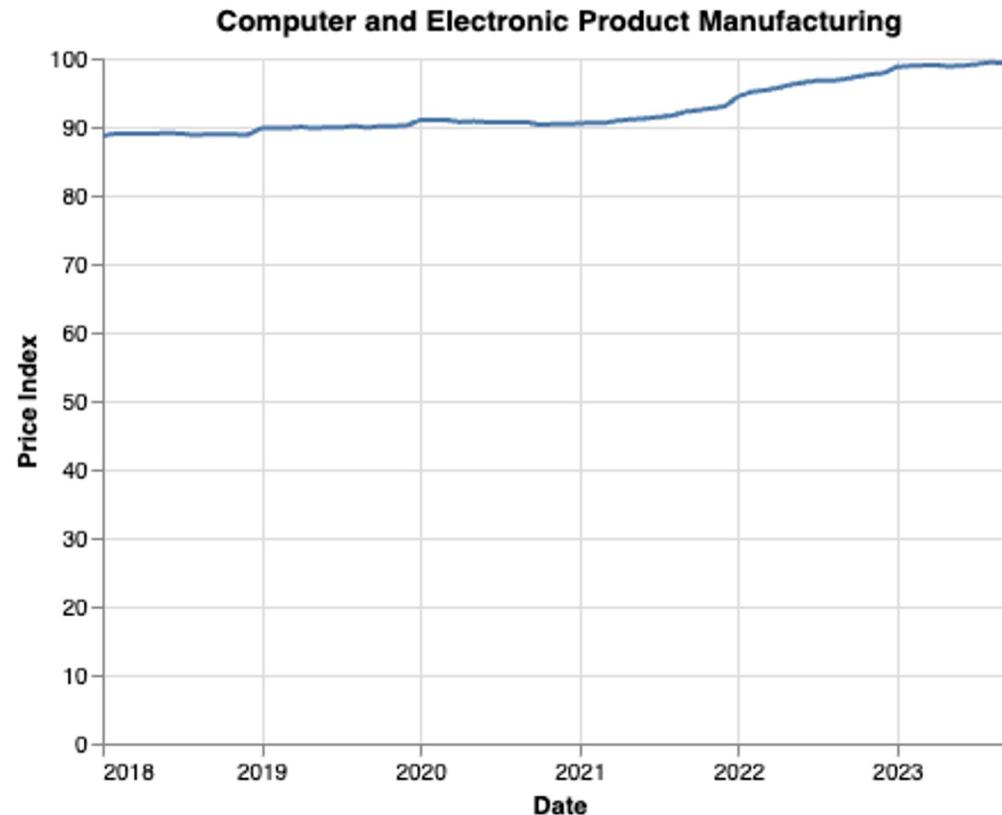
# Data Cleaning

- Mentions
  - Ensuring all mention counts were entire words
- Dates
  - Converted to DateTime format

# Keyword Mentions

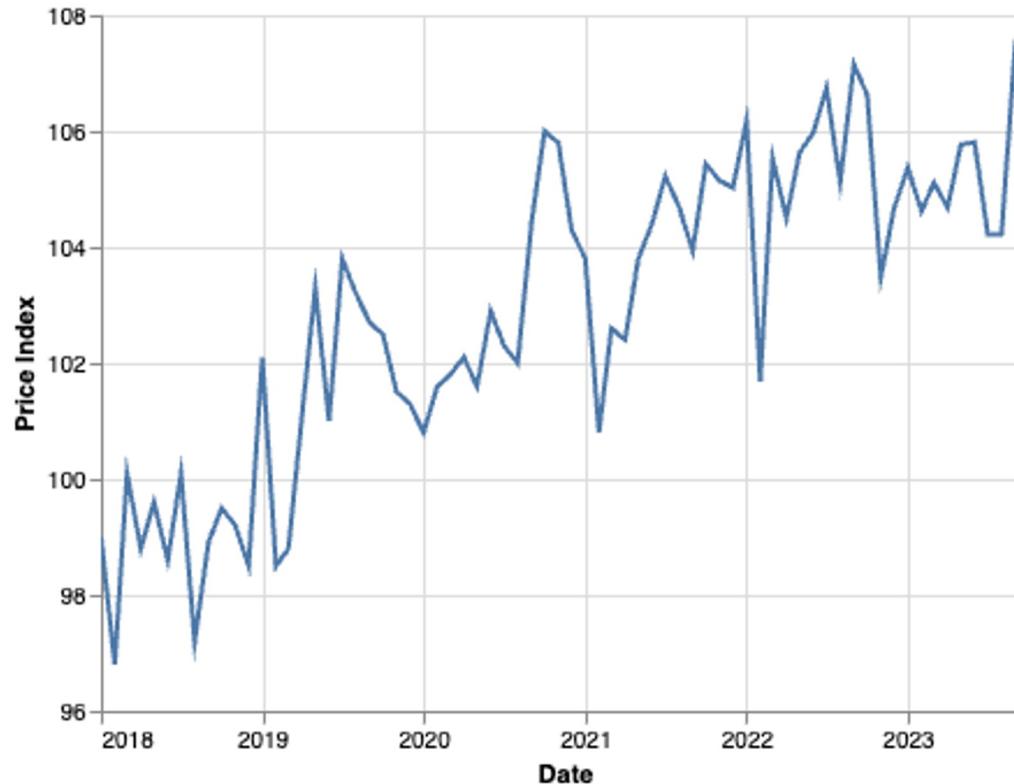
- Tracking two key phrases in transcripts
  - Artificial Intelligence (AI)
  - Machine Learning (ML)
  - Graphics Processing Unit (GPU)

# Computer and Electronic Product Manufacturing Industry Index (PCU)

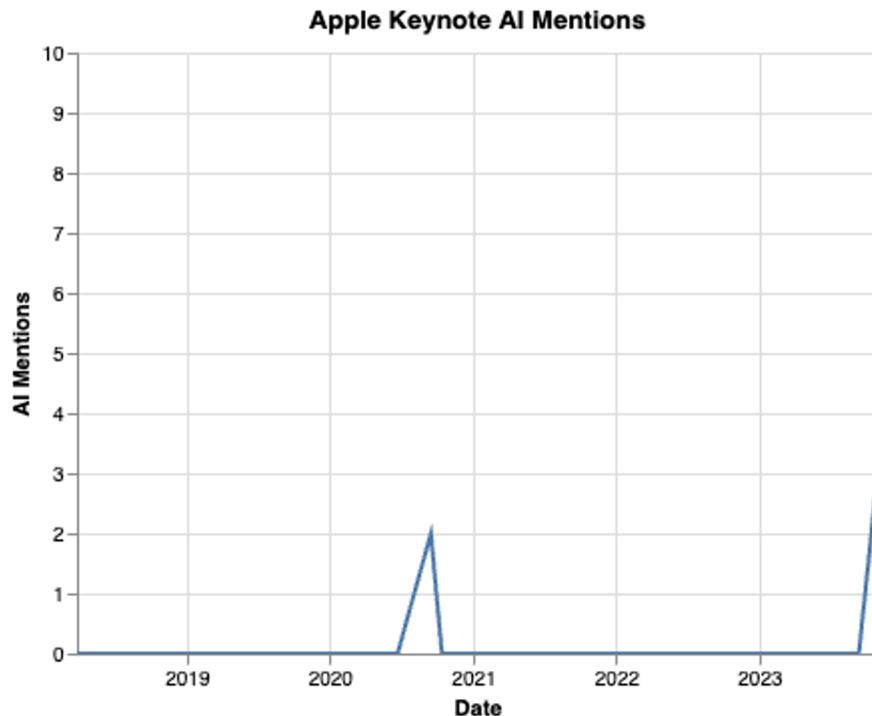
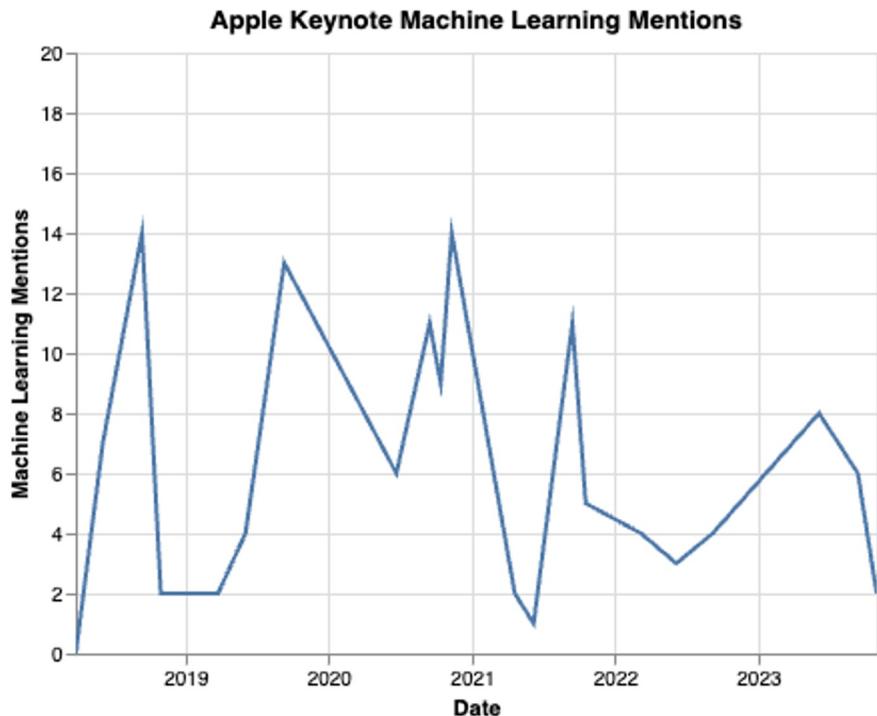


# Application Software Publishing Commodity Index (WPU)

Producer Price Index by Commodity: Software Publishing: Application Software Publishing

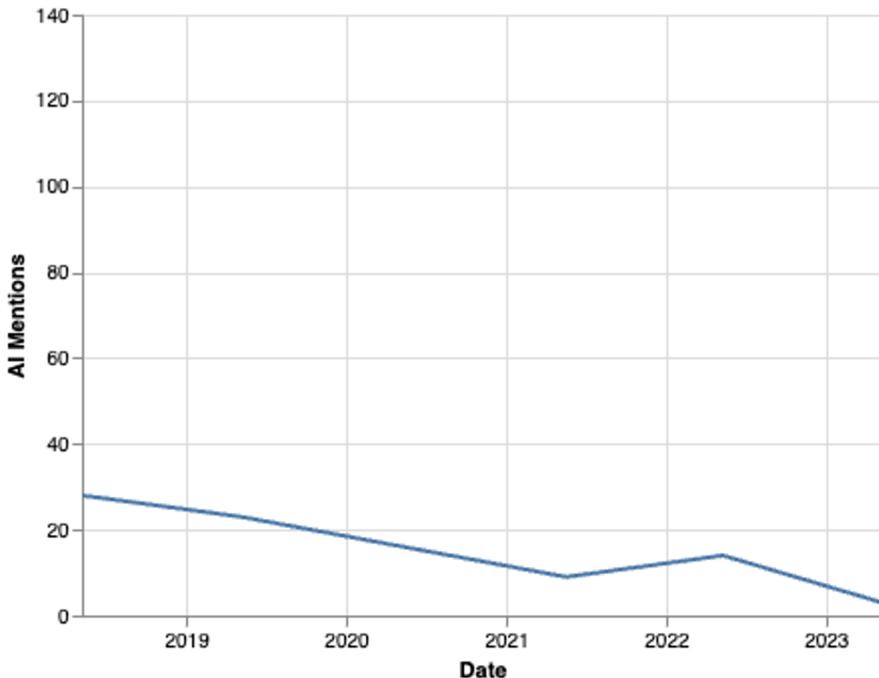


# Apple Mentions

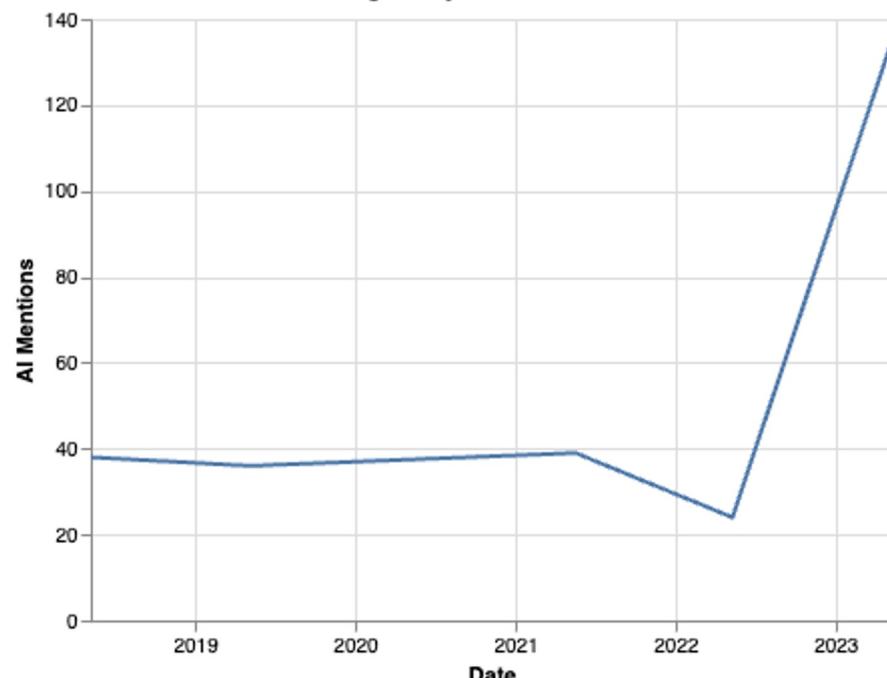


# Google

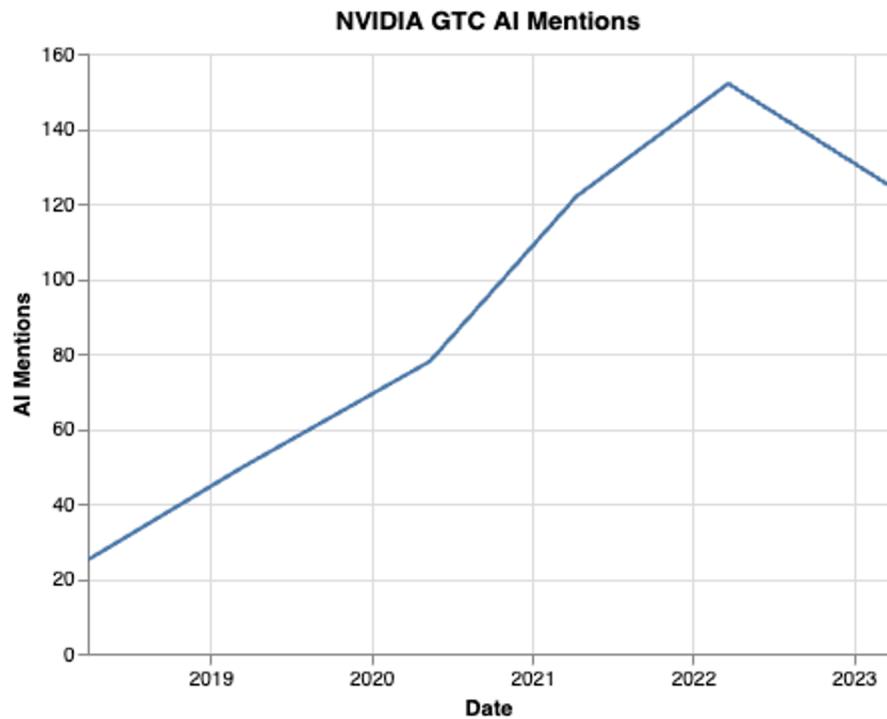
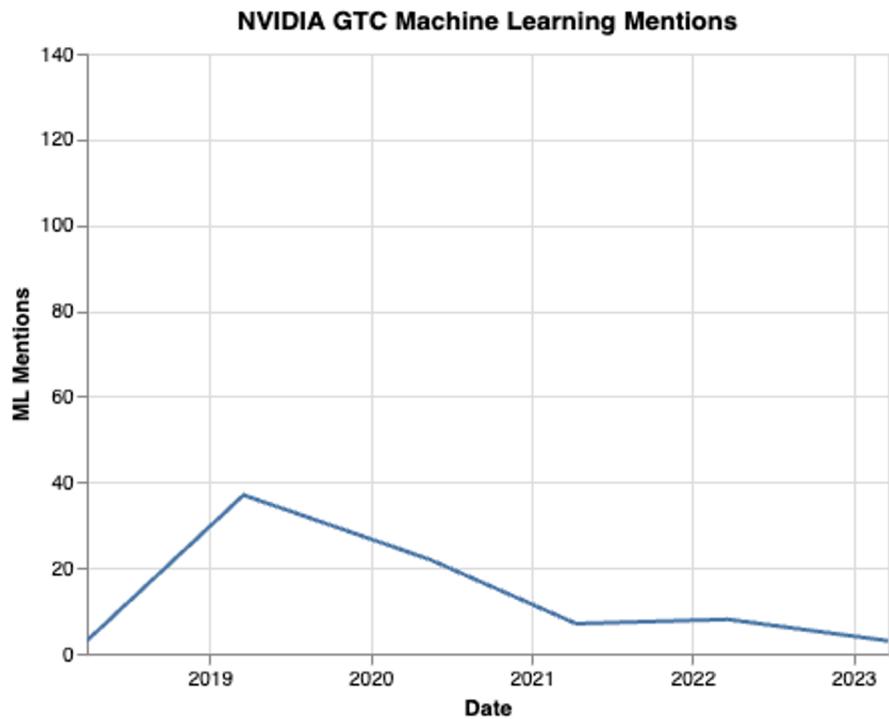
**Google Keynote Machine Learning Mentions**



**Google Keynote AI Mentions**



# Nvidia



# Index and Mention Collinearity

- Variance Inflation Factor (VIF)
  - VIF equal to 1: variables are not correlated
  - VIF between 1 and 5: variables are moderately correlated
  - VIF greater than 5: variables are highly correlated

$$\text{VIF}_i = \frac{1}{1 - R_i^2}$$

## VIF Index and Mentions

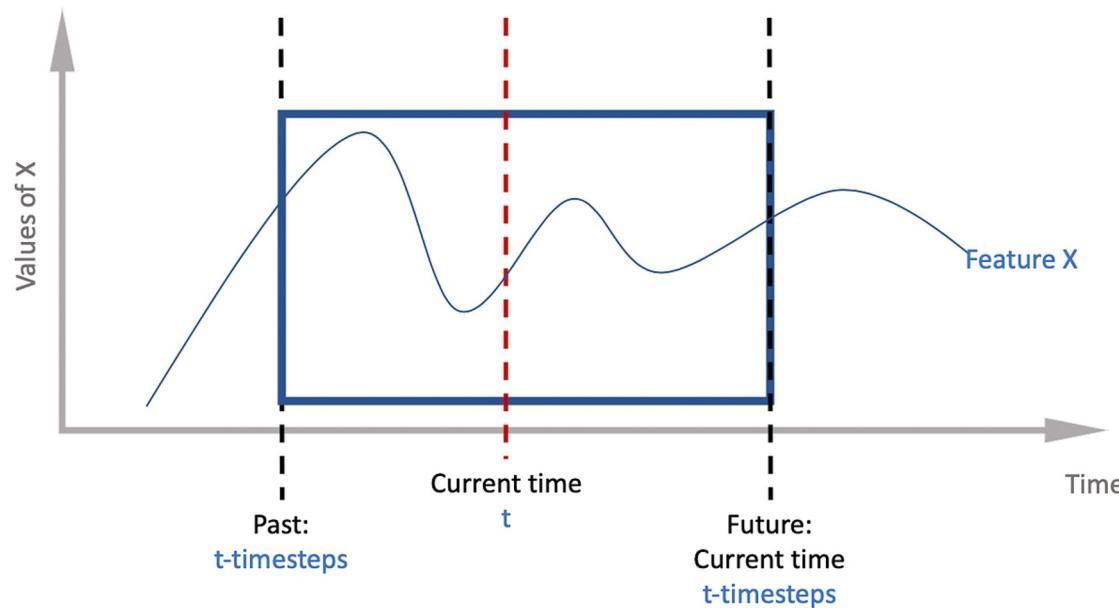
	WPU	PCU
Nvidia	5.714371	5.871394
Google	2.853679	3.007873
Apple	1.109174	1.109174

## VIF Index and Mentions

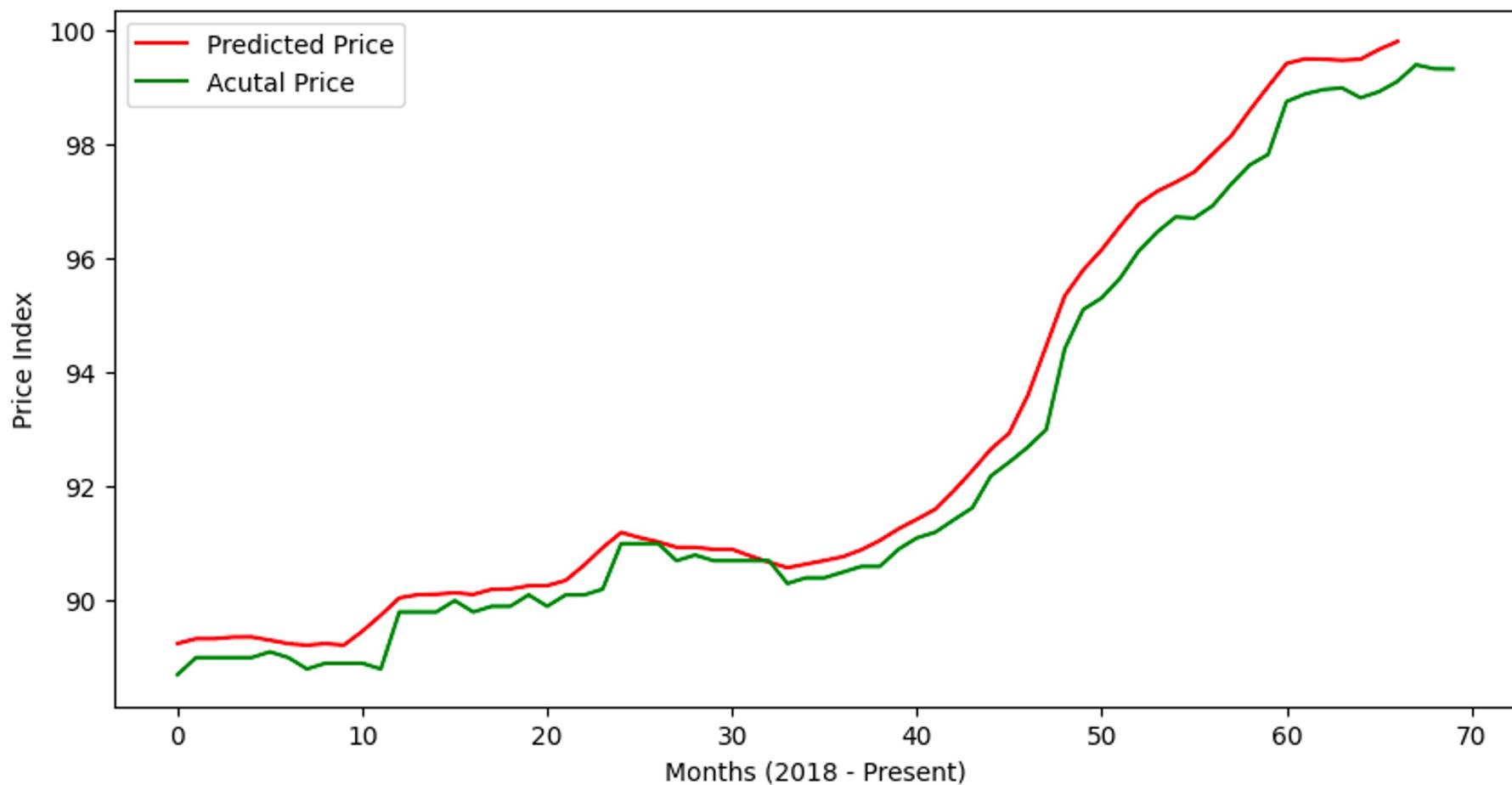
	WPU	PCU
Nvidia	5.714371	5.871394
Google	2.853679	3.007873
Apple	1.109174	1.109174

# Predicting Prices

- LSTM Model



## LSTM Prediction of PCU Index



## Projected Three Month Price

Projected Index Price: 100.48

Current Price: 99.33

% Change: 1.16%

Train RMSE: 0.0019611173775047064

# Letting the Data Speak

- We can use the same model that we used to predict price on words
- The model will look over our transcripts and tell us what word should come next

# 2018 Nvidia Transcript

Sentence: 'Our most truly important product is' -> Predicted next word: 'graphics'

	Date	Machine Learning	AI	GPU
4	2023-03-21		3	125
0	2022-03-22		8	152
1	2021-04-12		7	122
5	2020-05-14		22	78
2	2019-03-19		37	50
3	2018-03-28		3	25

# 2018 Nvidia Transcript

Sentence: 'Our most truly important product is' -> Predicted next word: 'graphics'

	Date	Machine Learning	AI	GPU
4	2023-03-21		3	125
0	2022-03-22		8	152
1	2021-04-12		7	122
5	2020-05-14		22	78
2	2019-03-19		37	50
3	2018-03-28		3	25

## Concluding Points

- Nvidia GTC drives the market when it comes to hardware commodities
- Google I/O has shifted its direction greatly towards AI adoption
- Apple is starting to dive back into AI to enhance their devices
- More data is needed to create model-based insight
- LSTM not the optimal choice
  - GPT and other transformers for self-attention

What types of ML tools are being developed?

## Question 2 Data

- **Artificial Intelligence Tools 2023**
  - **Variables:** AI Tool Name, Description, Free/Paid/Other, Useable For, Charges, Review, Tool Link, Major Category

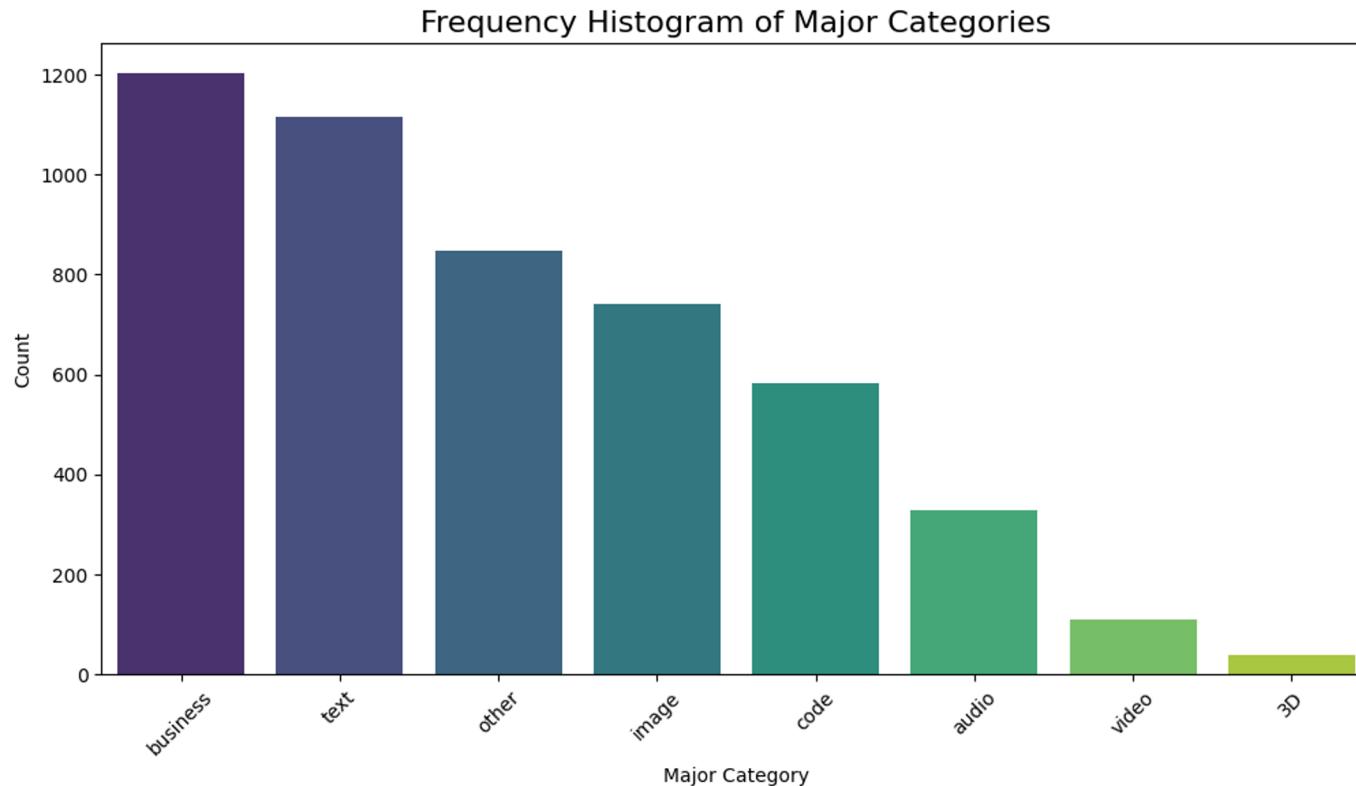
# Data Cleaning

- Dropping all missing values (from applicable features)
  - Dropping only values that will deter from the prediction

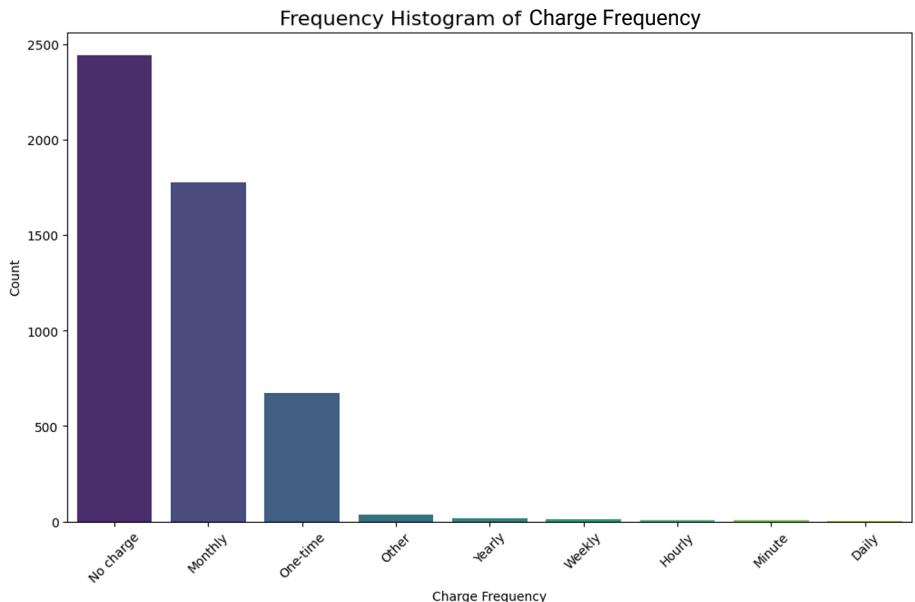
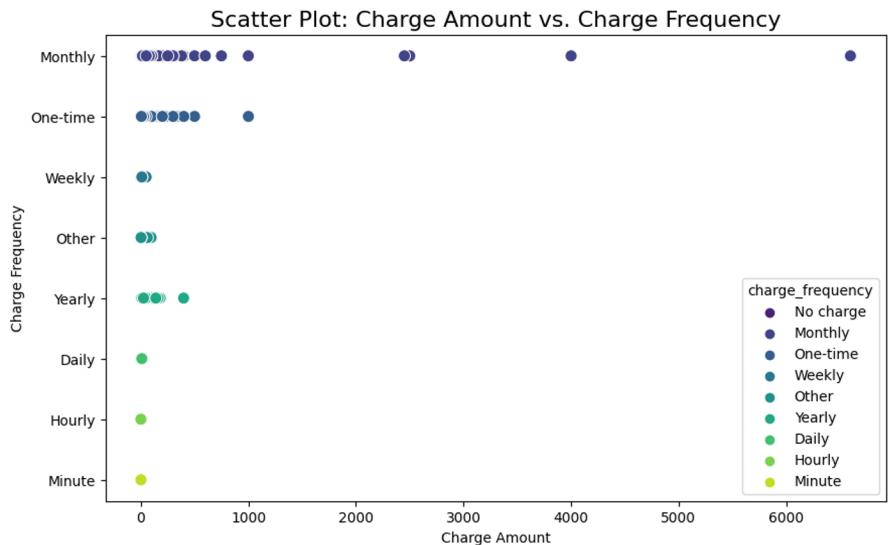
# Feature Engineering

- Charges
  - Two new variables
    - *charge\_frequency*
    - *charge\_amount*
- Input Missing Values for “Charge Amount”
  - All NaN values = 0

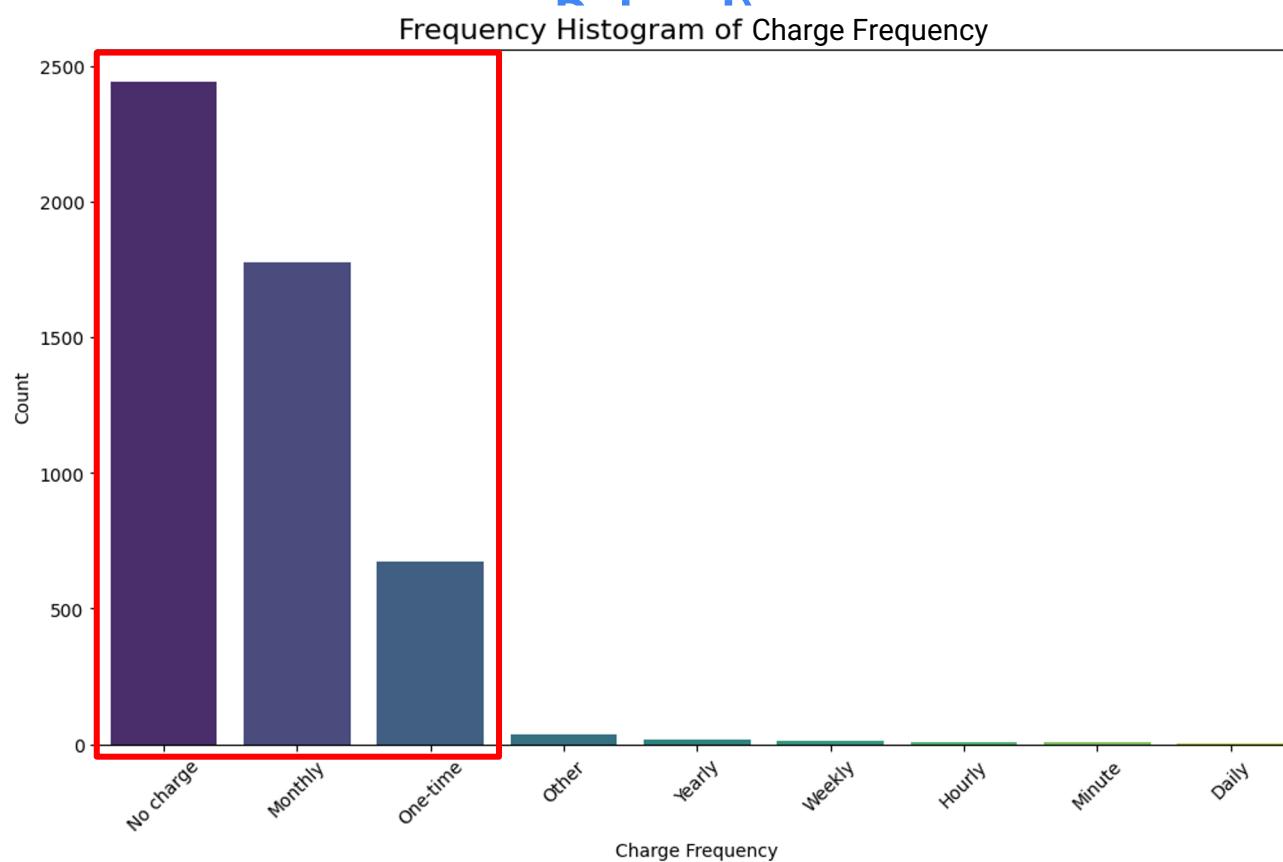
# AI Tools' Categories



# Costs & Its Frequencies

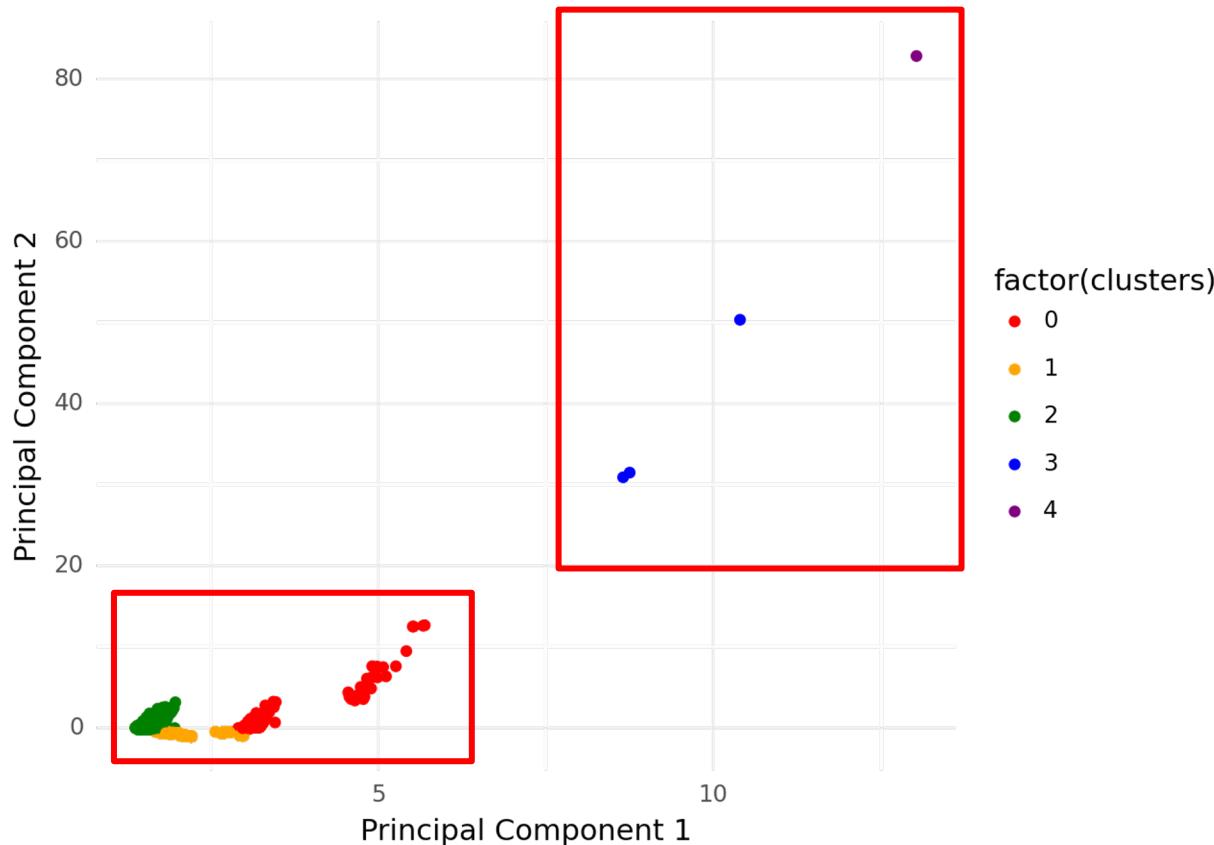


# AI Tools (Original Dataset) vs AI “Top Three” Tools (Subset)



# Clustering - KMeans

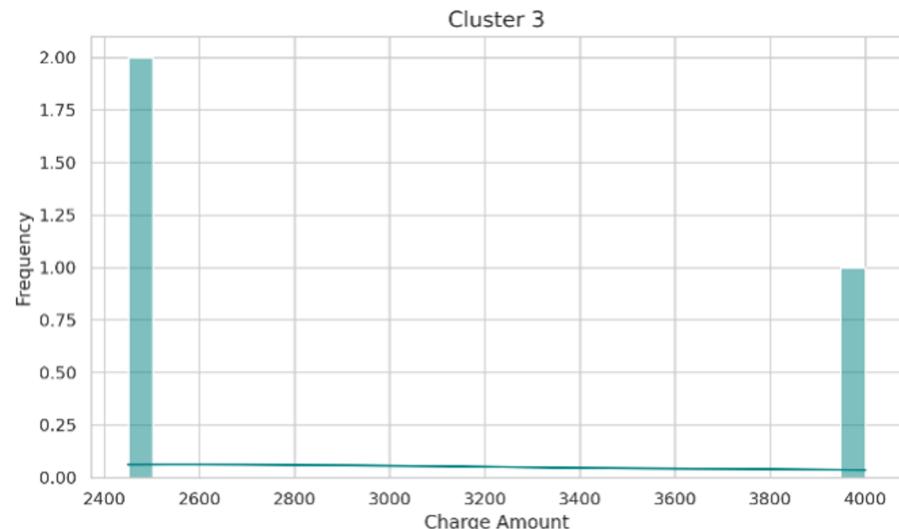
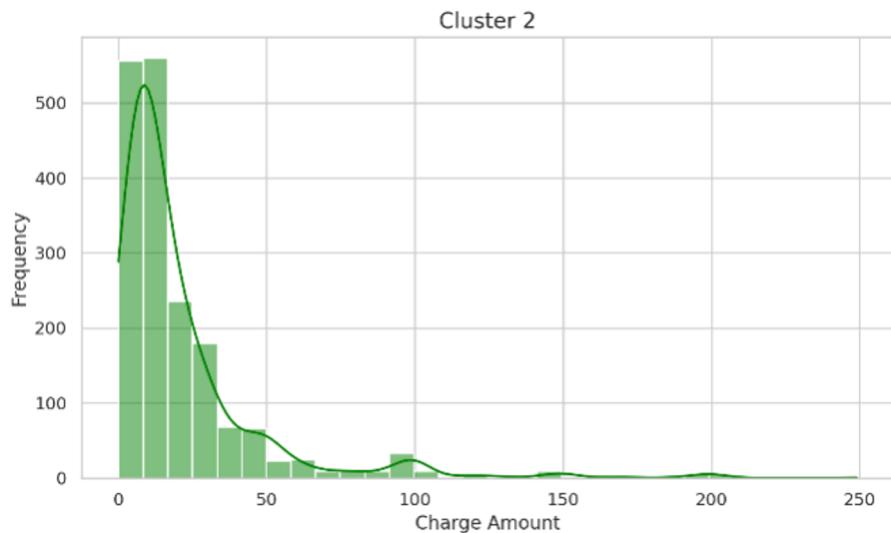
PCA Scatterplot of Clusters



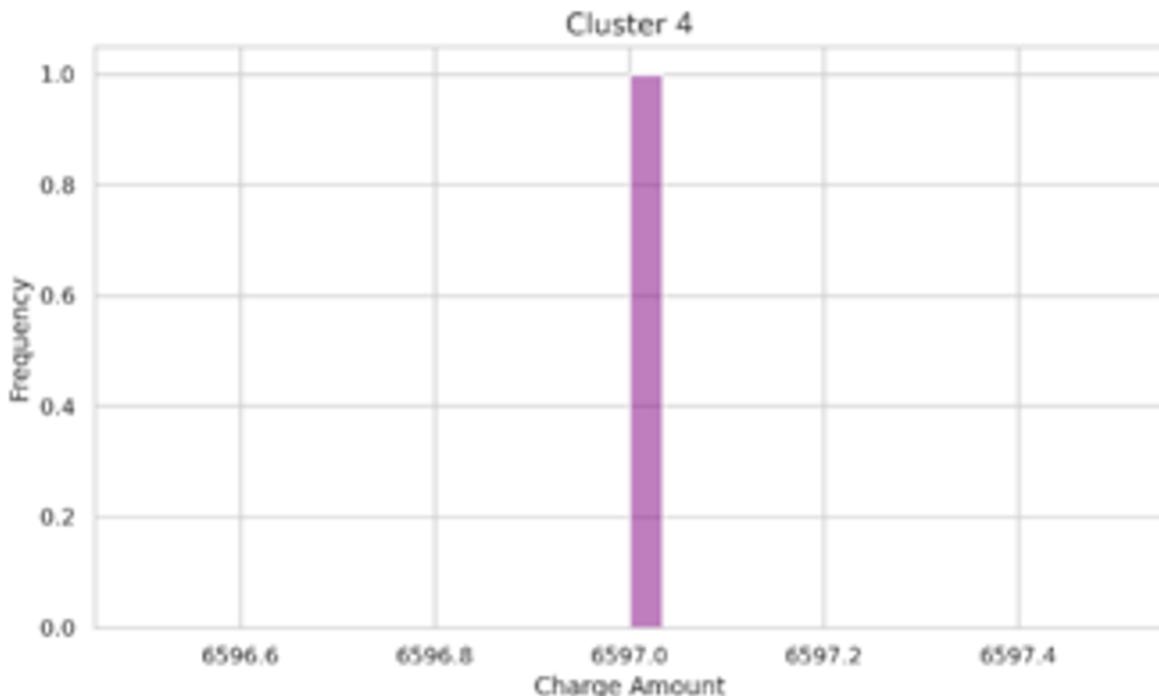
# “Charge Amount” Distribution Amongst Clusters 0 & 1



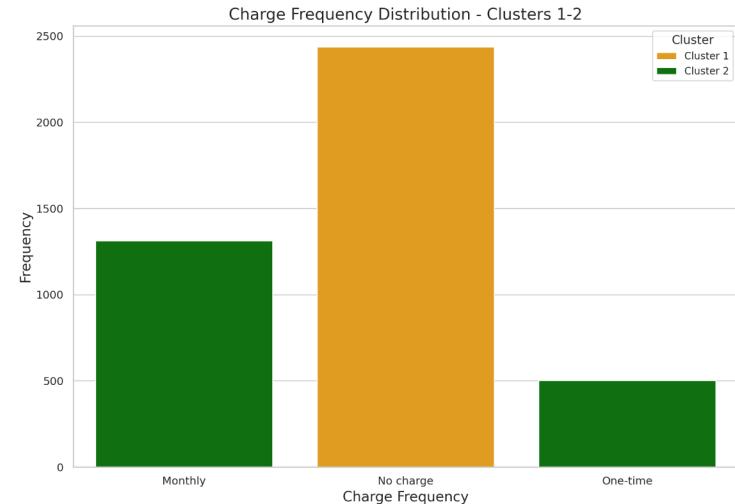
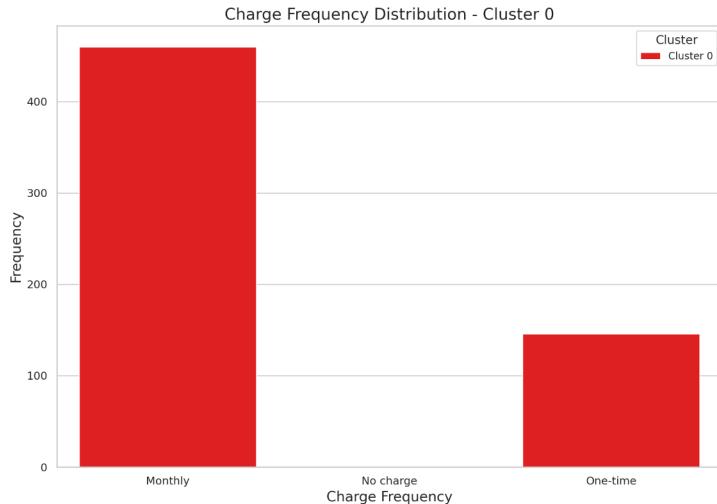
# “Charge Amount” Distribution Amongst Clusters 2 & 3



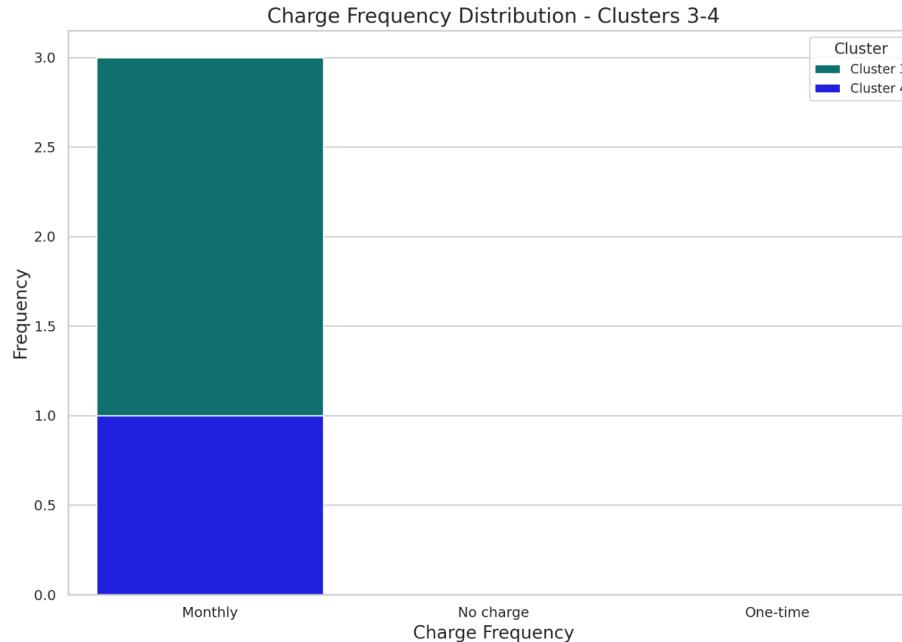
## “Charge Amount” Distribution Amongst Cluster 4



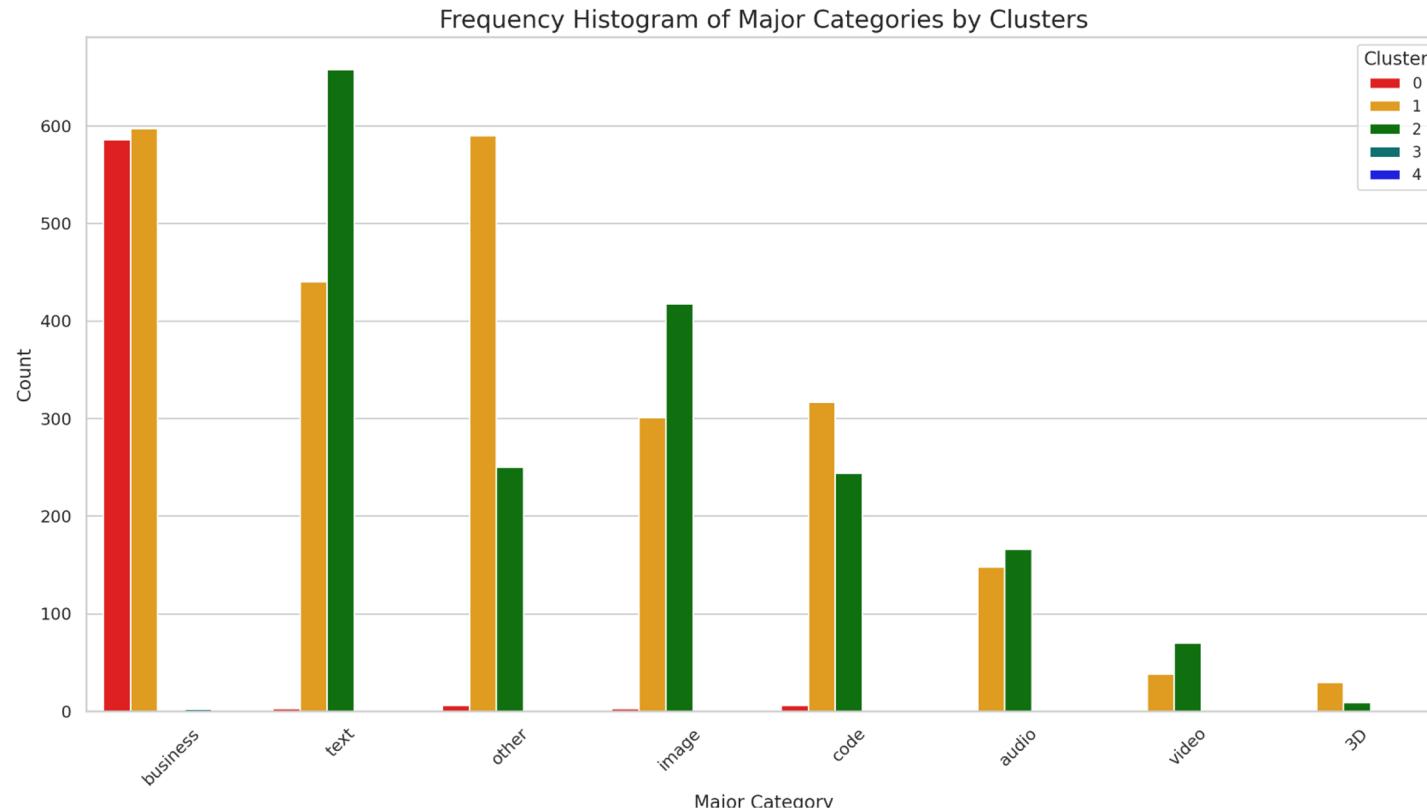
# “Charge Frequency” Distribution Amongst Clusters 0, 1, 2



# “Charge Frequency” Distribution Amongst Clusters 3, 4



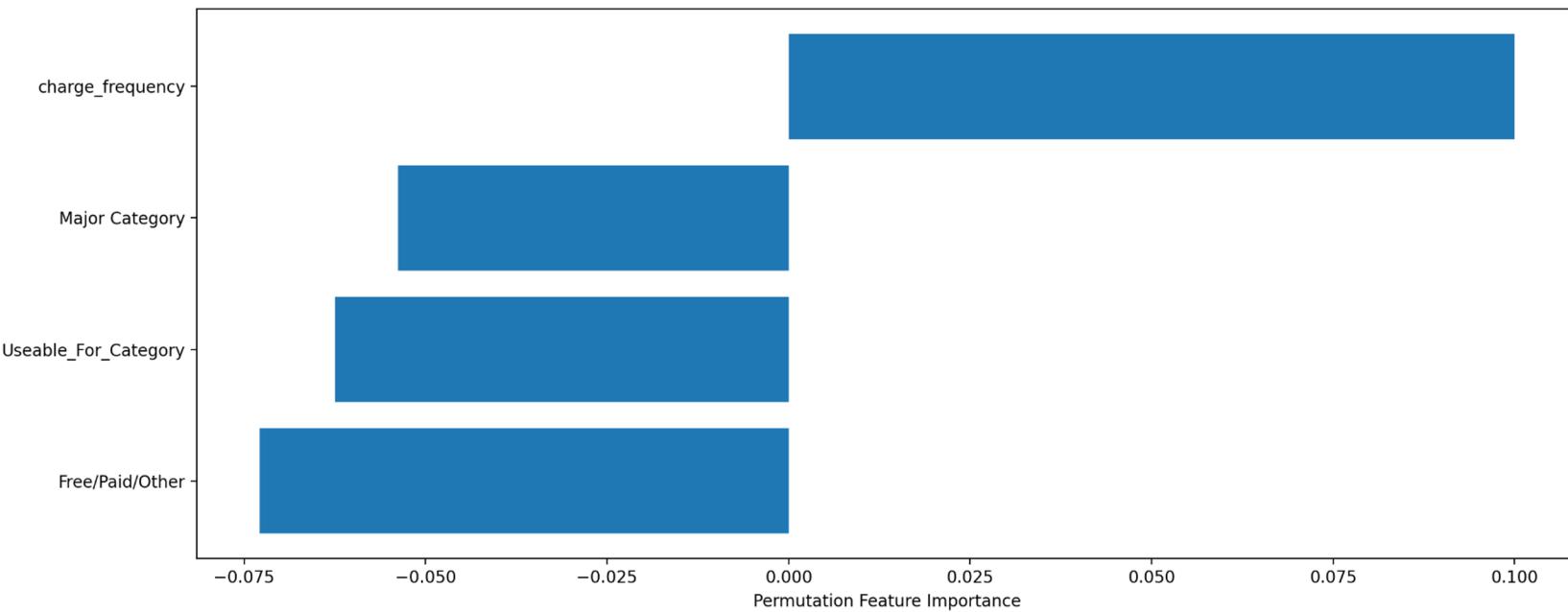
# “Major Categories” Distribution Amongst Clusters



# Predicting “Charge Amount”

- RandomForestRegressor() Model
- Performance Metrics
  - *Mean Squared Error (MSE)* ~15,280
    - How close our predictions are to the actual value
  - *Mean Absolute Error (MAE)* ~ 19.46
    - Average magnitude of errors
  - *Mean Absolute Percentage Error (MAPE)* ~826,006,337,922,890.6% means
    - Average % that our predictions deviate from the actual value
  - *R-squared (R2)* ~0.10
    - How much of the variability is explained

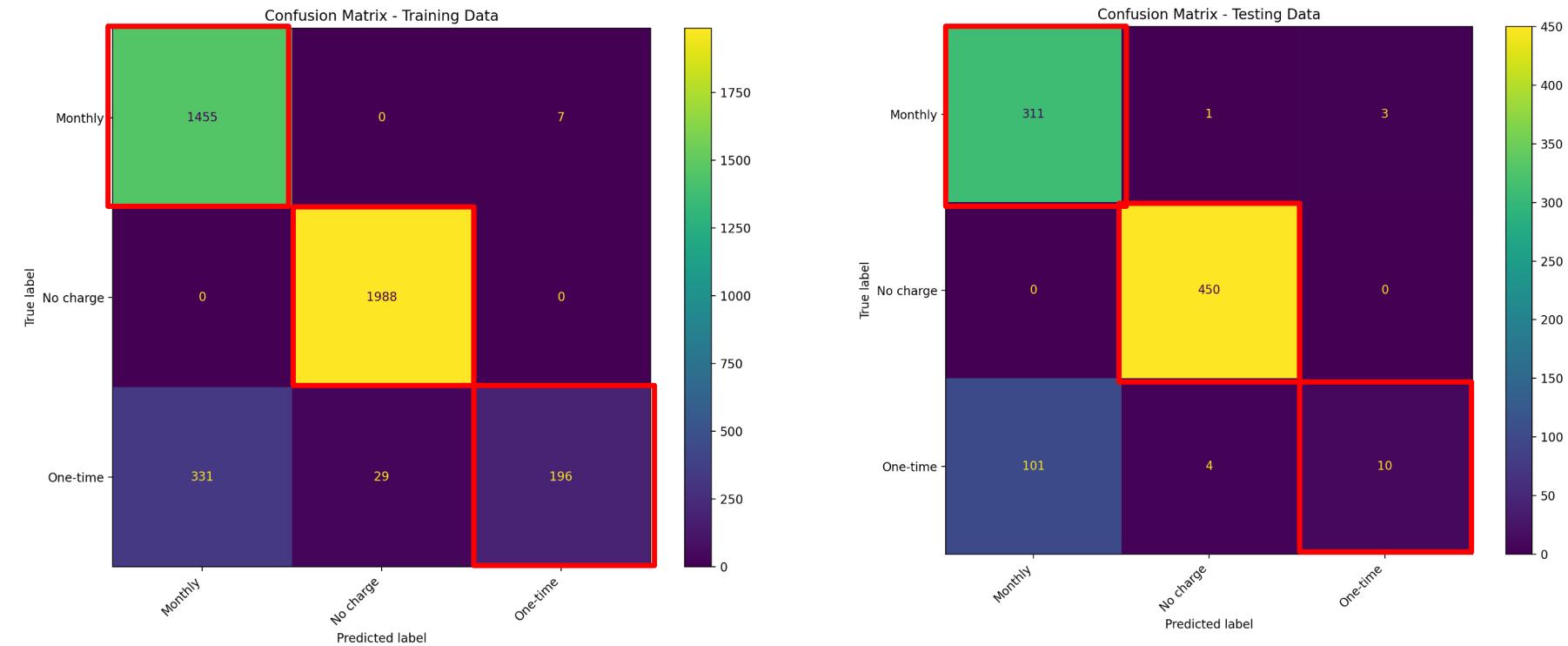
# Importance of the Features



# Classifying “Charge Frequency”

- RandomForestClassifier() Model
- Performance Metrics
  - *Train Acc* ~ 91%
    - How often is the model predicting correctly?
  - *Train Precision*: 92%
    - How many of the predicted positives are correct?
  - *Train Recall* ~91%
    - How often is the model correct for positive cases?
  - *Train F1* ~ 89%
    - Combination of the Recall and Precision score of a model

# “Charge Frequency” Prediction Accuracy



# Concluding Points on ML Tools

- **Builds off one another**
  - How they present themselves to the public
- **Designed to be Short Term, Low Commitment, Low Risk**
  - Introduce AI to a broader demographic
- **Minimize the decision making on consumers (renewing subscription plans)**
  - Increasing customer retention durations
- **Mostly made for business and text purposes**

**Thank you!**