

Data Science and Data Analytics

Prof. Dr. Huber, GroB (IBM 3)

Mira Heinemann and Esra Tonleu
400904676 and 400808321

Agenda

1. Defining the research question
2. Finding data
3. Setting up R
4. First attempt of coding
5. Sentiment Analysis Setup
6. Event-Window Approach
7. Current Status & Challenges
8. Next Steps
9. Feedback & Discussion

#1 Step

defining the research question



Die Wahrheit über Tesla

1,9 Mio. Aufrufe • vor 2 Monaten

 Simplicissimus ✓

Beim kostenlosen Konto von Revolut spart ihr durchschnittlich 120€ Kontoführungsgebühren. Zusätzlich könnt ihr euch bis zum ...

4K



100 MILLIARDEN DOLLAR WERT

Tesla überholt VW an der Börse

23.01.2020, 09:12

faz - Tesla überholt VW an der Börse (2020)

<https://www.youtube.com/watch?v=VuGFgHeaHNE>

≡ **Google Scholar** 

 **Artikel** Ungefähr 6.430 Ergebnisse (0,12 Sek.) 

Mein Profil  **Meine Bibliothek**

Beliebige Zeit
Seit 2025
Seit 2024
Seit 2021
Zeitraum wählen...

Nach Relevanz sortieren
Nach Datum sortieren

Beliebige Sprache
Seiten auf Deutsch

Alle Typen
Übersichtsarbeiten

Patente einschließen
 Zitate einschließen

 Alert erstellen

Tesla Inc. stock prediction using sentiment analysis
A Bhadkamar, S Bhattacharya - ... Accounting, Business and ..., 2022 - uwoajournals.org
... the unique relationship between **Elon Musk's Tweets** and **Tesla's stock** value. Exploratory ...
of **tweets/engagements** corresponded to an increase in **Tesla's** closing price and vice versa. ...
☆ Speichern  Zitiert von: 16 Ähnliche Artikel Alle 5 Versionen 

[PDF] uwoajournals.org
Full Text via CCOL

[PDF] ... impact of narcissistic CEOs running media companies on stock markets: A case study on Elon Musk's Twitter activity on the performance of Tesla and Twitter
L Huang - 2024 - ur.bc.edu
... an event study of **Elon Musk's Twitter** activities on **Tesla** and **Twitter** with a panel analysis ...
Elon Musk's tweets have price effects on **Tesla** and **Twitter's stock**, event studies on 48 **tweets** ...
☆ Speichern  Zitiert von: 1 Ähnliche Artikel 

Power of 280: Measuring the Impact of Elon Musk's Tweets on the Stock Market
S Metta, N Madhavan... - Ushus Journal of ..., 2022 - journals.christuniversity.in
... to which **Elon Musk's Twitter** posts may influence **stock** prices of ... , and returns obtained from listed **stocks**. Intra-day data from well... In this **tweet**, **Elon Musk** hints that **Tesla stock** prices are ...
☆ Speichern  Zitiert von: 11 Ähnliche Artikel Alle 3 Versionen 

CEO's tweets and firm stock returns: A case study of Elon Musk and Tesla
JG Dam - 2023 - digitalcommons.georgiasouthern ...
... from CEO **tweets** influences **Tesla's stock** returns. Because markets are inherently unstable, the interaction of fundaments and psychology may have nuanced effects on **Tesla** returns ...
☆ Speichern  Zitiert von: 4 Ähnliche Artikel Alle 2 Versionen 

Sentiment analysis of Elon Musk's Twitter data using LSTM and ANFIS-SVM
B Erkatal, A Yilmaz - International Conference on Intelligent and Fuzzy ..., 2022 - Springer
... The **Tesla's** and SpaceX's CEO and owner **Elon Musk's tweets** had a huge impact on coin ...
... inflance of **Elon Musk's tweets** on economic indexes Bitcoin (BTC), Dogecoin & **Tesla Stock**. ...
☆ Speichern  Zitiert von: 6 Ähnliche Artikel Alle 2 Versionen 

Impact of Twitter on Stock Market
JS Šević, N Stakić, AJ Stakić - Proceedings of the 1st ..., 2023 - books.google.com

[PDF] atlantis-press.com
ejournals@cambridge

Sentiment analysis of **Elon Musk's Twitter** data using LSTM and ANFIS-SVM

B Erkatal, A Yilmaz - International Conference on Intelligent and Fuzzy ..., 2022 - Springer

... The **Tesla**'s and SpaceX's CEO and owner **Elon Musk**'s **tweets** had a huge impact on coin ... inflance of **Elon Musk**'s **tweets** on economic indexes Bitcoin (BTC), Dogecoin & **Tesla Stock** ...

☆ Speichern 99 Zitieren Zitiert von: 6 Ähnliche Artikel Alle 2 Versionen ☺

Impact of Twitter on Stock Market

JS Šević, N Stakić, AJ Stakić - Proceedings of the 1st ..., 2023 - books.google.com

... between **Elon Musk**'s **Twitter** activities and **stock** price performance of **Tesla** company, by ... In this paper we analyze the relationship between **Musk**'s **tweets** and **Tesla stock** price, with ...

☆ Speichern 99 Zitieren Zitiert von: 3 Ähnliche Artikel Alle 2 Versionen

[PDF] atlantis-press.com

ejournals@cambridge

The effect of social media activity on the market: case **Elon Musk**, evidence from the US

J Kuokka - 2021 - lutpub.lut.fi

... Keywords: **Elon Musk**, crypto, **Twitter**, **Tesla** With the rise of **Twitter** almost all the known ... on the **stock** or crypto market. The full data of this study consists of **Elon Musk**'s **tweets** from the ...

☆ Speichern 99 Zitieren Zitiert von: 2 Ähnliche Artikel ☺

[PDF] lut.fi

Predicting **Tesla Stock** Return Using **Twitter** Data-An Intraday View on the Relation between **Twitter** Dimensions and the **Tesla Stock** Return

G Edman, M Weishaupt - 2020 - lup.lub.lu.se

... The only difference is that we do not include a dummy for **Elon Musk**, as none of his **tweets** includes "\$**TSLA**" in the observed period. We present the results for the one-minute interval in ...

☆ Speichern 99 Zitieren Ähnliche Artikel ☺

[PDF] lu.se

[PDF] **Tesla**: Is Now the Time to Invest? An examination of **Tesla**, social media, and its effect on **stock**

M Coiro - ... of **Tesla**, social media, and its effect on **stock** (August ..., 2021 - researchgate.net

... **Tesla**'s Chief Executive, **Elon Musk** has made bold statements ... of **tweets** and **stock** price fluctuations in order to gain clarity and understanding on how posts on **Twitter** associate to **stock** ...

☆ Speichern 99 Zitieren Zitiert von: 2 Ähnliche Artikel Alle 2 Versionen ☺

[PDF] researchgate.net

[PDF] THE TWEETING 500: CAN WE PREDICT THE IMPACT ON THE STOCK MARKET?

T Monsch - 2021 - arno.uvt.nl

... , recent **stock** market rallies partly fueled by **Tesla**'s CEO **Elon Musk** ... Thus, we expect to find that **tweets** influence **stock** price ... **Elon Musk**, we have seen immediate effects of the **tweets** on ...

☆ Speichern 99 Zitieren Ähnliche Artikel ☺

[PDF] uvt.nl

“What is the effect of Elon Musk’s tweets on Tesla’s stock market performance?”

Why is the question economically relevant?

- Social media usage of CEO's can influence financial markets
- Investors react within minutes to public statements
- Elon Musk's tweets have repeatedly affected Tesla's stock price and trading volume
- Understanding these reactions helps investors, companies, and regulators manage risks

#2 Step

finding data

← elon musk tweet

Comments 262 Notebooks 140 Datasets 42 Topics 6

Filter by

CREATOR

- You 0
- Others 42

42 Results

Relevance ▾

Elon Musk Tweets
Dataset · 2y ago · by Gabriel Preda
Tweets of Elon Musk collected using tweepy

Elon Musk's Tweets Dataset 2022
Dataset · 3y ago · by Marta Castrillo
Context > Elon Musk is a business magnate and investor.

Elon Musk Tweets (Daily Updated)
Dataset · 2y ago · by Aryan Singh
Tweets of Elon Musk collected using snscreape.

Elon Musk's Tweets
Dataset · 8y ago · by E. Kaan Ulgen
Context Elon Musk is an American business magnate.

All Elon Musk's Tweets
Dataset · 5y ago · by Andrada
2010 - 2021 tweets by Elon Musk

Elon Musk Tweets Dataset (17K)
Dataset · 3y ago · by Yasir Raza
Dataset of Elon Musk tweets till now (17K)

+ Create

Home

Competitions

Datasets

Models

Benchmarks

Game Arena

<> Code

Discussions

Learn

More

View Active Events



ALEX HUGGLER · UPDATED 3 YEARS AGO

29

<> Code

Download



Elon Musk Tweets Sentiment(Classified via RoBERTa)

Elon Musk Historical Twitter Data



Data Card Code (5) Discussion (0) Suggestions (0)

About Dataset

Usability ⓘ

10.00

License

CC0: Public Domain

Expected update frequency

Quarterly

Tags

News

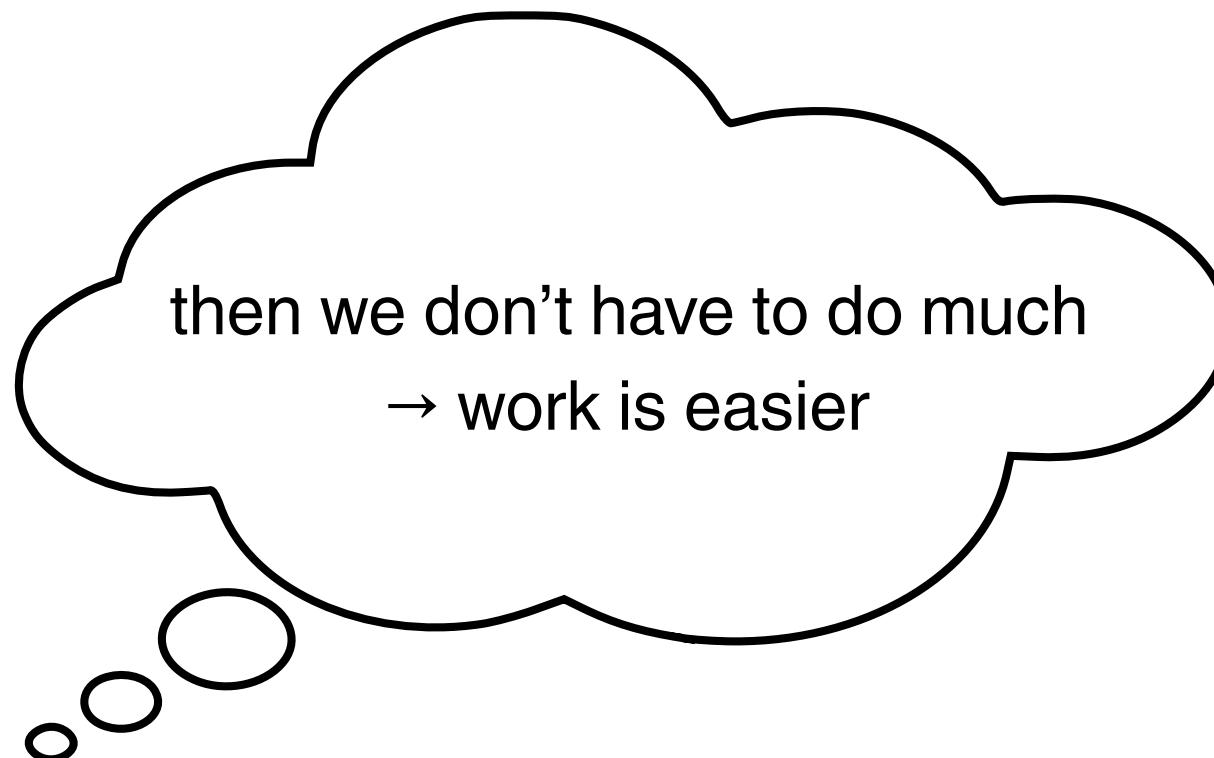
Automobiles and Vehicles

Electronics Electricity

Popular Culture

This dataset was compiled by scraping Twitter for tweets from the [@elonmusk](#) account. In addition, feature engineering has occurred on the dataset to add a field called 'sentiment' and 'mentions', for further analysis. The sentiment column was generated by using a RoBERTa Transformers Model, which can be found here (<https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment>). This dataset can be used to show time series analysis of [@elonmusk](#) tweets, can be used for other forms of data analysis, text analysis, sentiment analysis, and other various forms of NLP. **Update: Newly added files will be added with the most recent refresh date listed in the file name, current update is as of 12-25-22**

This dataset was compiled by scraping Twitter for tweets from the [@elonmusk](#) account. In addition, feature engineering has occurred on the dataset to add a field called 'sentiment' and 'mentions', for further analysis. The sentiment column was generated by using a RoBERTa Transformers Model, which can be found here (<https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment>). This dataset can be used to show time series analysis of [@elonmusk](#) tweets, can be used for other forms of data analysis, text analysis, sentiment analysis, and other various forms of NLP. **Update: Newly added files will be added with the most recent refresh date listed in the file name, current update is as of 12-25-22**





Elon Musk Tweet Dataset (Kaggle)

Contains tweet text, timestamps, and engagement metrics
(tweets, retweets, replies, quotes)



Tesla Stock Data from Yahoo

Includes daily closing price, returns, and trading volume



R's quantmod package downloads data directly from the Yahoo Finance API, which provides clean financial data, daily updates, easy access via one command.

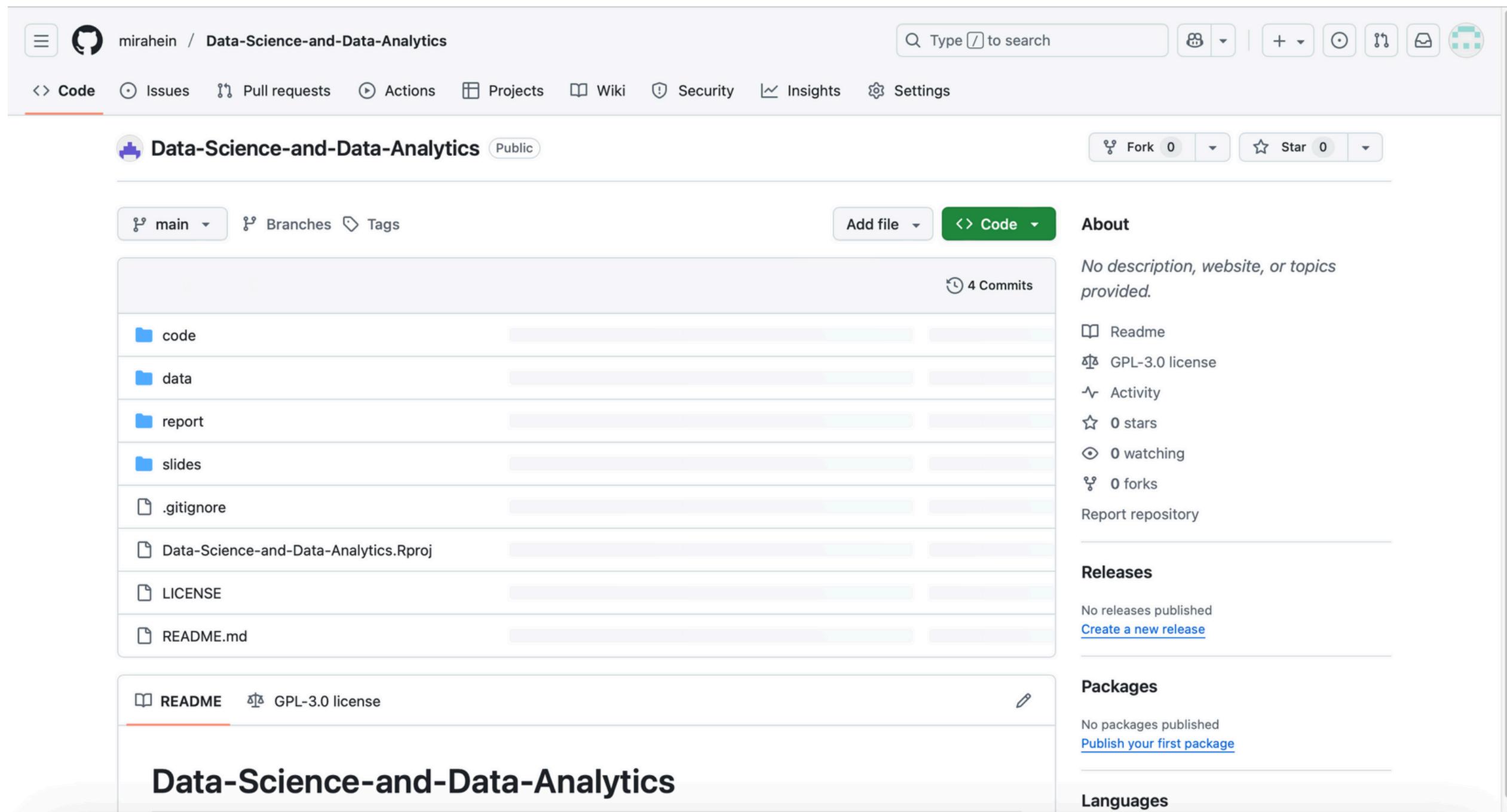
#3 Step

setting up R

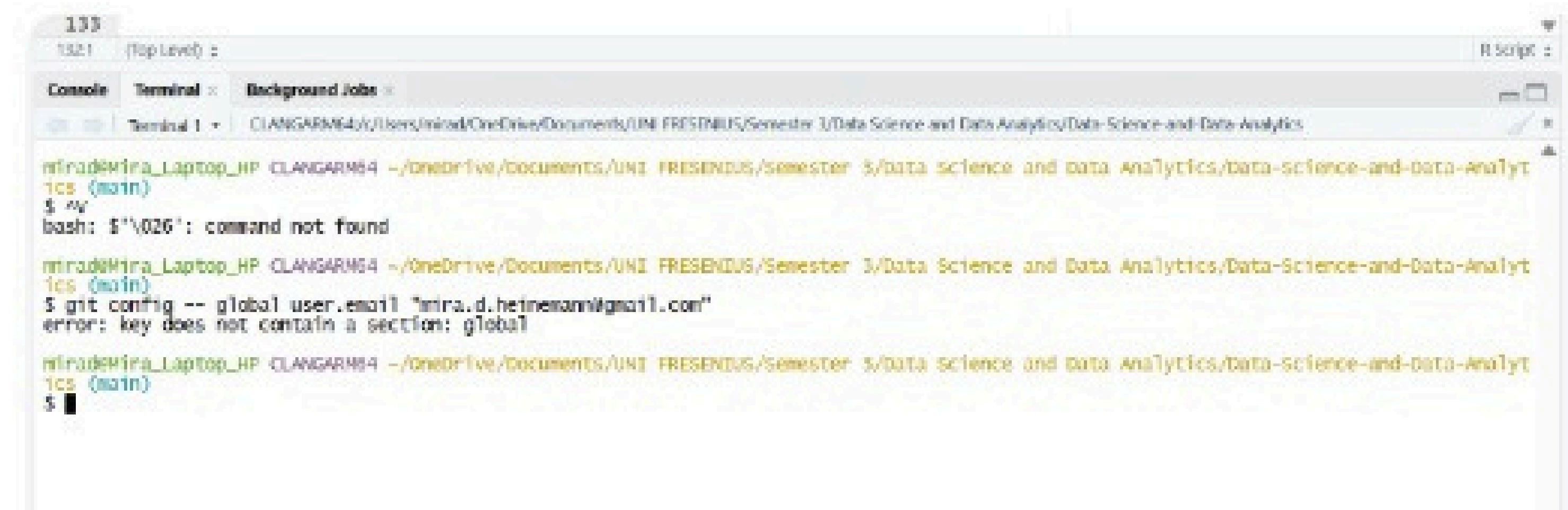
Premises

1. working on the same project simultaneously without overwriting
2. structure

1. we created a Shared Github Repository and named *Data-Science-and-Data-Analytics*



2. we cloned a project in R and linked it to GitHub
3. we set up Push-Commit-Pull



The screenshot shows the RStudio interface with a terminal window open. The terminal tab is selected, showing the following session:

```
137
1321 (Top Level) R Script

Console Terminal x Background Jobs
Terminal 1 ~ CLANGARMS4:/Users/Unire/OneDrive/Documents/UNI FRESENIUS/Semester 1/Data Science and Data Analytics/Data-Science-and-Data-Analyt
ics (main)
$ R
bash: $'\026': command not found

miradewira_Laptop_HP CLANGARMS4 ~/OneDrive/Documents/UNI FRESENIUS/Semester 1/Data Science and Data Analytics/Data-Science-and-Data-Analyt
ics (main)
$ git config -- global user.email "mirad.heinemann@gmail.com"
error: key does not contain a section: global

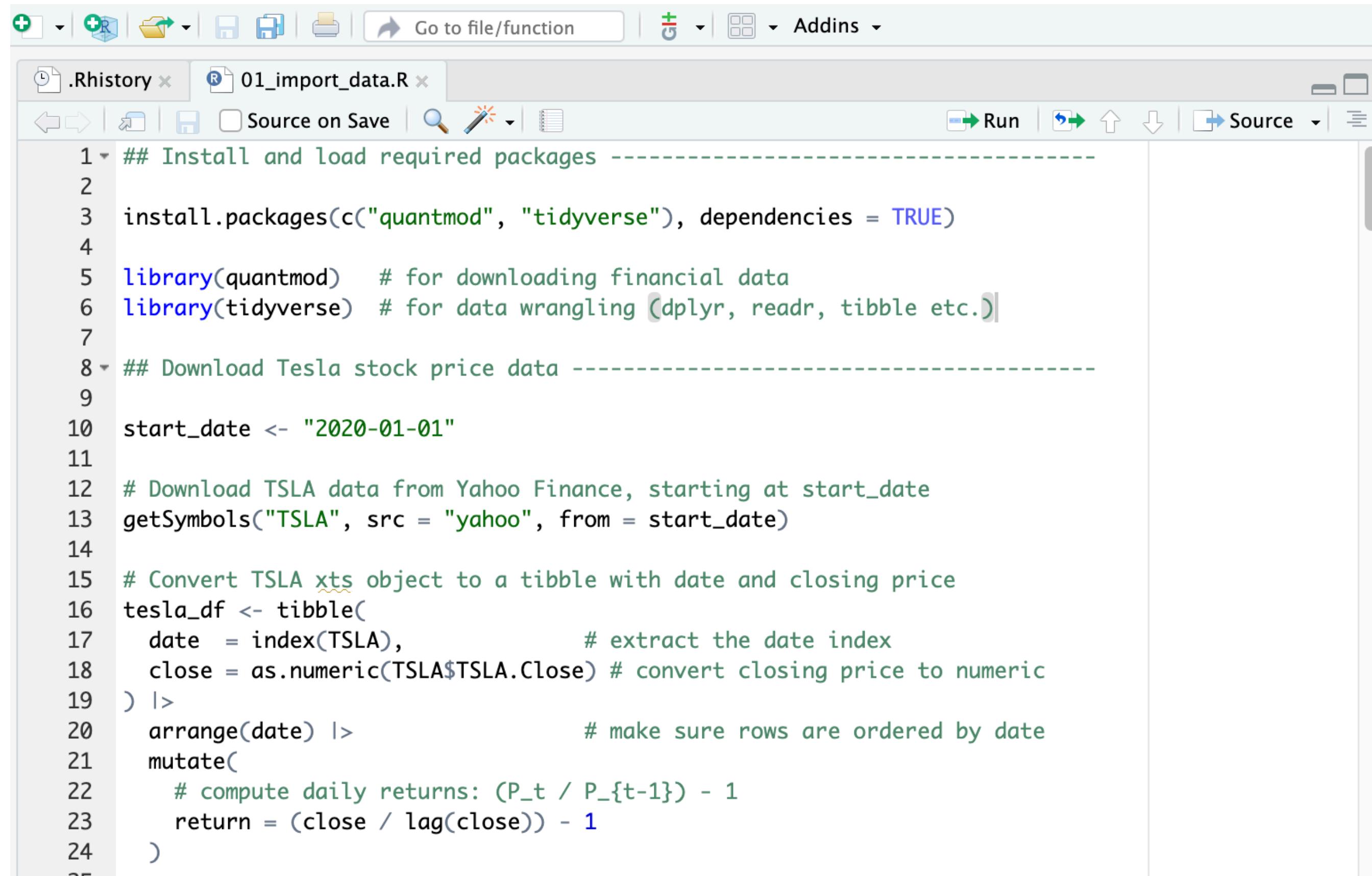
miradewira_Laptop_HP CLANGARMS4 ~/OneDrive/Documents/UNI FRESENIUS/Semester 1/Data Science and Data Analytics/Data-Science-and-Data-Analyt
ics (main)
$
```

4.

```
dir.create("data")
dir.create("code")
dir.create("figs")
dir.create("slides")
dir.create("report")
```

#4 Step

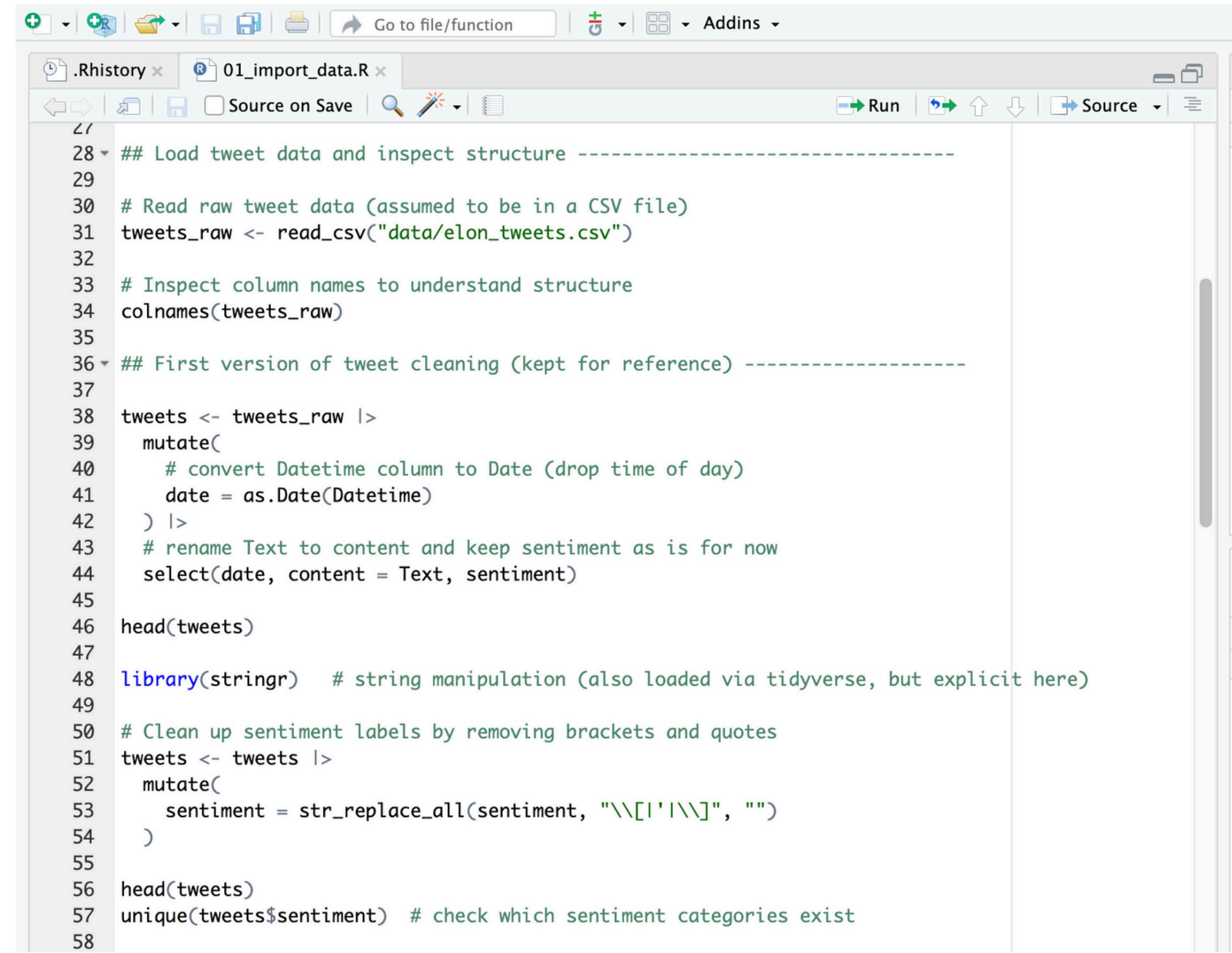
first attempt of coding



The screenshot shows the RStudio interface with the following details:

- Toolbar:** Includes icons for file operations (New, Open, Save, Print, Find, Go to file/function, Addins).
- Tab Bar:** Shows ".Rhistory" and "01_import_data.R" as the active file.
- Toolbar:** Includes icons for Source on Save, Find, Run, Source, and other navigation.
- Code Editor:** Displays the R script content.

```
1 ## Install and load required packages -----
2
3 install.packages(c("quantmod", "tidyverse"), dependencies = TRUE)
4
5 library(quantmod) # for downloading financial data
6 library(tidyverse) # for data wrangling (dplyr, readr, tibble etc.)
7
8 ## Download Tesla stock price data -----
9
10 start_date <- "2020-01-01"
11
12 # Download TSLA data from Yahoo Finance, starting at start_date
13 getSymbols("TSLA", src = "yahoo", from = start_date)
14
15 # Convert TSLA xts object to a tibble with date and closing price
16 tesla_df <- tibble(
17   date = index(TSLA), # extract the date index
18   close = as.numeric(TSLA$TSLA.Close) # convert closing price to numeric
19 ) |>
20   arrange(date) # make sure rows are ordered by date
21   mutate(
22     # compute daily returns: (P_t / P_{t-1}) - 1
23     return = (close / lag(close)) - 1
24   )
25
```



The screenshot shows the RStudio interface with the following details:

- File Bar:** Contains icons for file operations like Open, Save, and Print, along with a "Go to file/function" search bar and an "Addins" dropdown.
- Project Bar:** Shows ".Rhistory" and "01_import_data.R" as the active files.
- Toolbar:** Includes buttons for "Source on Save", "Run", "Source", and other common functions.
- Code Editor:** Displays the R script content. The script is for importing tweet data from a CSV file and cleaning it using the tidyverse and stringr packages.

```
27
28 ## Load tweet data and inspect structure -----
29
30 # Read raw tweet data (assumed to be in a CSV file)
31 tweets_raw <- read_csv("data/elon_tweets.csv")
32
33 # Inspect column names to understand structure
34 colnames(tweets_raw)
35
36 ## First version of tweet cleaning (kept for reference) -----
37
38 tweets <- tweets_raw |>
39   mutate(
40     # convert Datetime column to Date (drop time of day)
41     date = as.Date(Datetime)
42   ) |>
43   # rename Text to content and keep sentiment as is for now
44   select(date, content = Text, sentiment)
45
46 head(tweets)
47
48 library(stringr) # string manipulation (also loaded via tidyverse, but explicit here)
49
50 # Clean up sentiment labels by removing brackets and quotes
51 tweets <- tweets |>
52   mutate(
53     sentiment = str_replace_all(sentiment, "\\"[!`!\\]", "")
54   )
55
56 head(tweets)
57 unique(tweets$sentiment) # check which sentiment categories exist
58
```

```
--  
59 ## Aggregate tweets by day -----  
60  
61 tweets_daily <- tweets |>  
62   group_by(date) |>  
63   summarise(  
64     tweet_count = n(),                      # number of tweets per day  
65     pos = sum(sentiment == "positive"),       # count of positive tweets  
66     neu = sum(sentiment == "neutral"),         # count of neutral tweets  
67     neg = sum(sentiment == "negative"),         # count of negative tweets  
68     .groups = "drop"  
69   )  
70
```

```

--> 89 ## Alternative (equivalent) aggregation and merge -----
--> 90
--> 91 tweets_daily <- tweets %>%
--> 92   group_by(date) %>%
--> 93   summarise(
--> 94     tweet_count = n(),
--> 95     pos = sum(sentiment == "positive"),
--> 96     neu = sum(sentiment == "neutral"),
--> 97     neg = sum(sentiment == "negative"),
--> 98     .groups = "drop"
--> 99   )
--> 100
--> 101 data_merged <- tesla_df %>%
--> 102   left_join(tweets_daily, by = "date") %>%
--> 103   # here we use replace_na() with a named list to fill NAs
--> 104   replace_na(list(
--> 105     tweet_count = 0,
--> 106     pos = 0,
--> 107     neu = 0,
--> 108     neg = 0
--> 109   )) %>%
--> 110   mutate(tweet_day = if_else(tweet_count > 0, 1L, 0L))
--> 111
--> 112 head(data_merged)
--> 113

--> 135
--> 136 ## Re-aggregate tweets by day with cleaned sentiment -----
--> 137
--> 138 unique(tweets$sentiment)
--> 139
--> 140 tweets_daily <- tweets |>
--> 141   group_by(date) |>
--> 142   summarise(
--> 143     tweet_count = n(),           # total tweets per day
--> 144     pos = sum(sentiment == "positive"), # positive tweets per day
--> 145     neu = sum(sentiment == "neutral"), # neutral tweets per day
--> 146     neg = sum(sentiment == "negative"), # negative tweets per day
--> 147     .groups = "drop"
--> 148   )
--> 149
--> 150 head(tweets_daily)
--> 151

```

Limiting our analysis

Several conceptual challenges

- Multiple tweets on the same day
- Difficulty visualizing patterns over long periods

DECISION: Limit our analysis to around 10 tweets and conduct a focused event-study

Minute interval vs daily Stock Market

≡ **kaggle**

Search

MOHAMMED ZAHID · UPDATED 4 YEARS AGO

▲ 2 ◀ ▶ Code Download ⋮

Tesla intraday stock market data

5 minute interval data from Jan 2017- June 2020



Data Card Code (0) Discussion (0) Suggestions (0)

About Dataset

This dataset contains open, high, low, closing price as well as volume data(OHLCV) for 5 min intervals

Usability 3.53

License Unknown

Expected update frequency Not specified

Tags

Finance Investing

≡ **kaggle**

Search

MOHAMMED ZAHID · UPDATED 4 YEARS AGO

▲ 2 ◀ ▶ Code Download ⋮

Tesla intraday stock market data

5 minute interval data from Jan 2017- June 2020



Data Card Code (0) Discussion (0) Suggestions (0)

About Dataset

This dataset contains open, high, low, closing price as well as volume data(OHLCV) for 5 min intervals

Usability 3.53

License Unknown

Expected update frequency Not specified

Tags

Finance Investing

≡ **kaggle**

Search

MOHAMMED ZAHID · UPDATED 4 YEARS AGO

▲ 2 ◀ ▶ Code Download ⋮

Tesla intraday stock market data

5 minute interval data from Jan 2017- June 2020



Data Card Code (0) Discussion (0) Suggestions (0)

About Dataset

This dataset contains open, high, low, closing price as well as volume data(OHLCV) for 5 min intervals

Usability 3.53

License Unknown

Expected update frequency Not specified

Tags

Finance Investing

≡ **kaggle**

Search

MOHAMMED ZAHID · UPDATED 4 YEARS AGO

▲ 2 ◀ ▶ Code Download ⋮

Tesla intraday stock market data

5 minute interval data from Jan 2017- June 2020



Data Card Code (0) Discussion (0) Suggestions (0)

About Dataset

This dataset contains open, high, low, closing price as well as volume data(OHLCV) for 5 min intervals

Usability 3.53

License Unknown

Expected update frequency Not specified

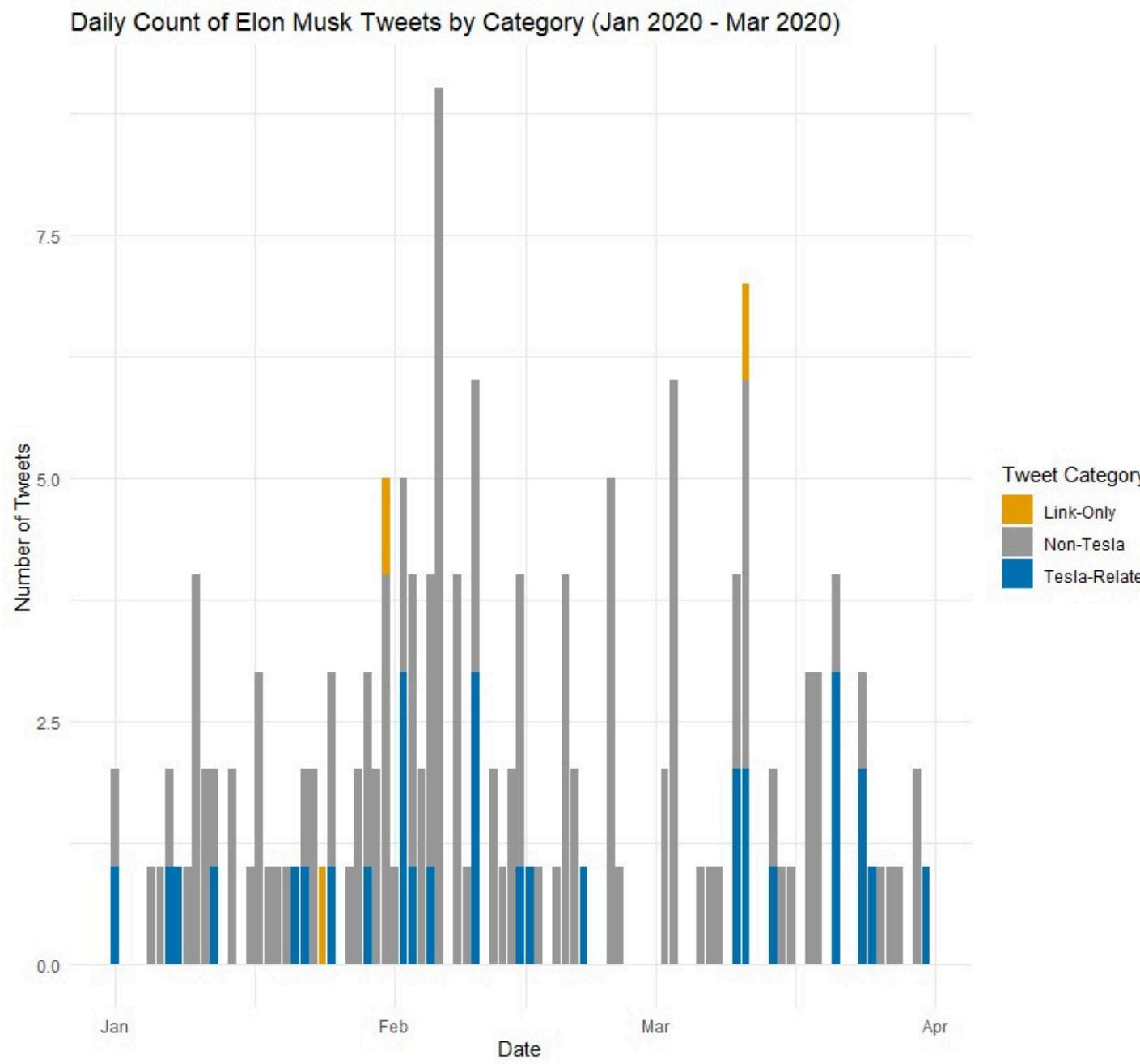
Tags

Finance Investing

Filtering for Original Tweets

To ensure we analyse only Musk's
communication we removed:
Retweets, Replies, Quote Tweets

Only using Tesla related Tweets

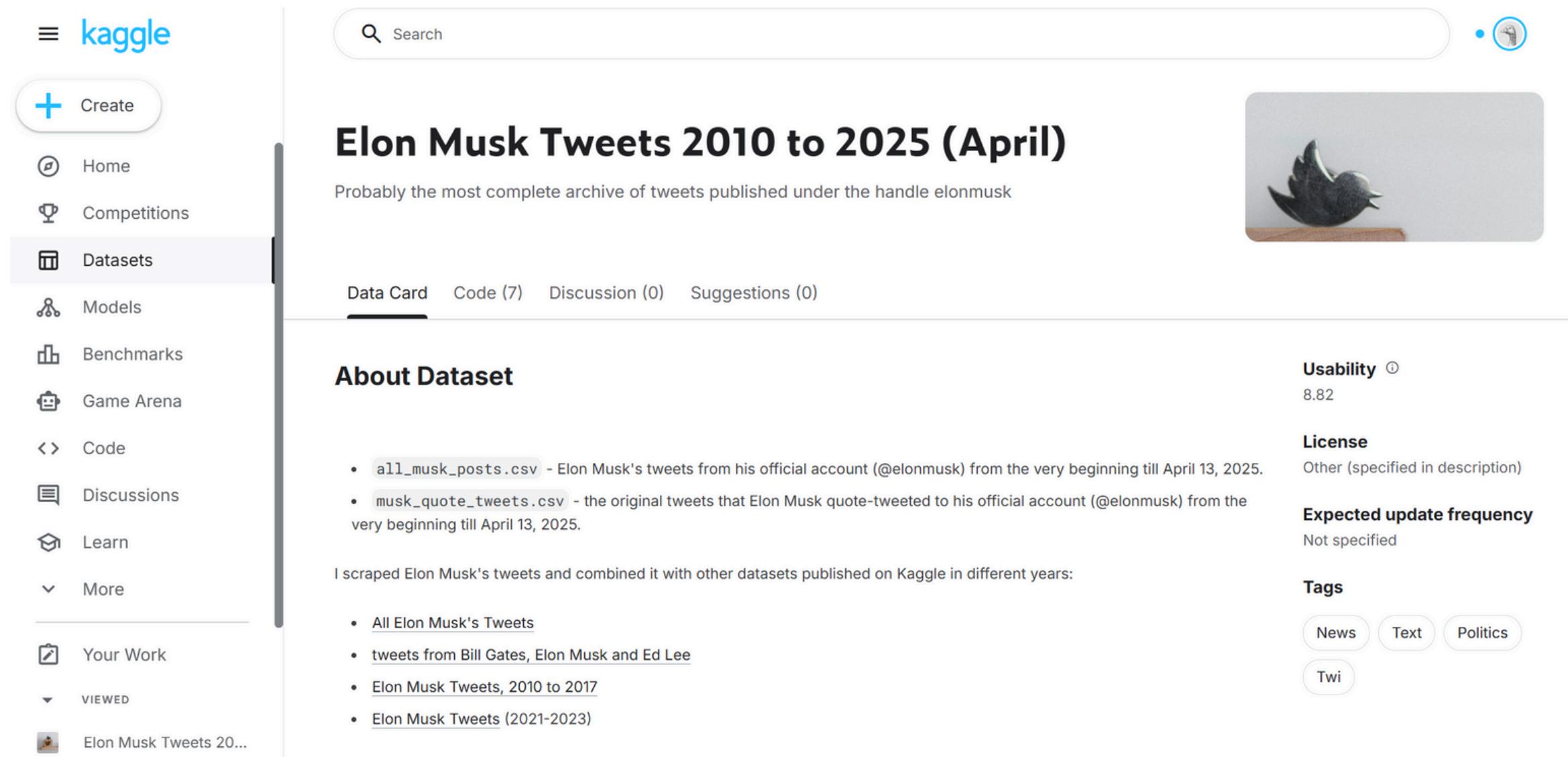


tesla keywords:

"tesla", "tsla",
"model s", "model 3", "model x", "model y",
"cybertruck", "roadster",
"fsd", "full self-driving", "autopilot",
"gigafactory", "giga berlin", "giga shanghai", "giga texas",
"supercharger", "megapack", "powerwall"

Switch Tweet Data Set

Unsure how reliable the sentiment labels in the original tweet dataset were



The screenshot shows the Kaggle interface. The left sidebar has a 'Datasets' tab selected. The main page displays a dataset titled 'Elon Musk Tweets 2010 to 2025 (April)'. The description states, 'Probably the most complete archive of tweets published under the handle elonmusk'. Below the title is a small image of a Twitter bird. The page includes sections for 'About Dataset', 'Data Card', 'Code (7)', 'Discussion (0)', and 'Suggestions (0)'. The 'About Dataset' section lists two CSV files: 'all_musk_posts.csv' and 'musk_quote_tweets.csv'. It also notes that the dataset was scraped from Kaggle. The 'Data Card' section shows a 'Usability' rating of 8.82, a 'License' of 'Other (specified in description)', and an 'Expected update frequency' of 'Not specified'. The 'Tags' section includes 'News', 'Text', 'Politics', and 'Twi'.

Elon Musk Tweets 2010 to 2025 (April)

Probably the most complete archive of tweets published under the handle elonmusk

Data Card Code (7) Discussion (0) Suggestions (0)

About Dataset

- [all_musk_posts.csv](#) - Elon Musk's tweets from his official account (@elonmusk) from the very beginning till April 13, 2025.
- [musk_quote_tweets.csv](#) - the original tweets that Elon Musk quote-tweeted to his official account (@elonmusk) from the very beginning till April 13, 2025.

I scraped Elon Musk's tweets and combined it with other datasets published on Kaggle in different years:

- [All Elon Musk's Tweets](#)
- [tweets from Bill Gates, Elon Musk and Ed Lee](#)
- [Elon Musk Tweets, 2010 to 2017](#)
- [Elon Musk Tweets \(2021-2023\)](#)

Usability 8.82

License Other (specified in description)

Expected update frequency Not specified

Tags

News Text Politics

Twi

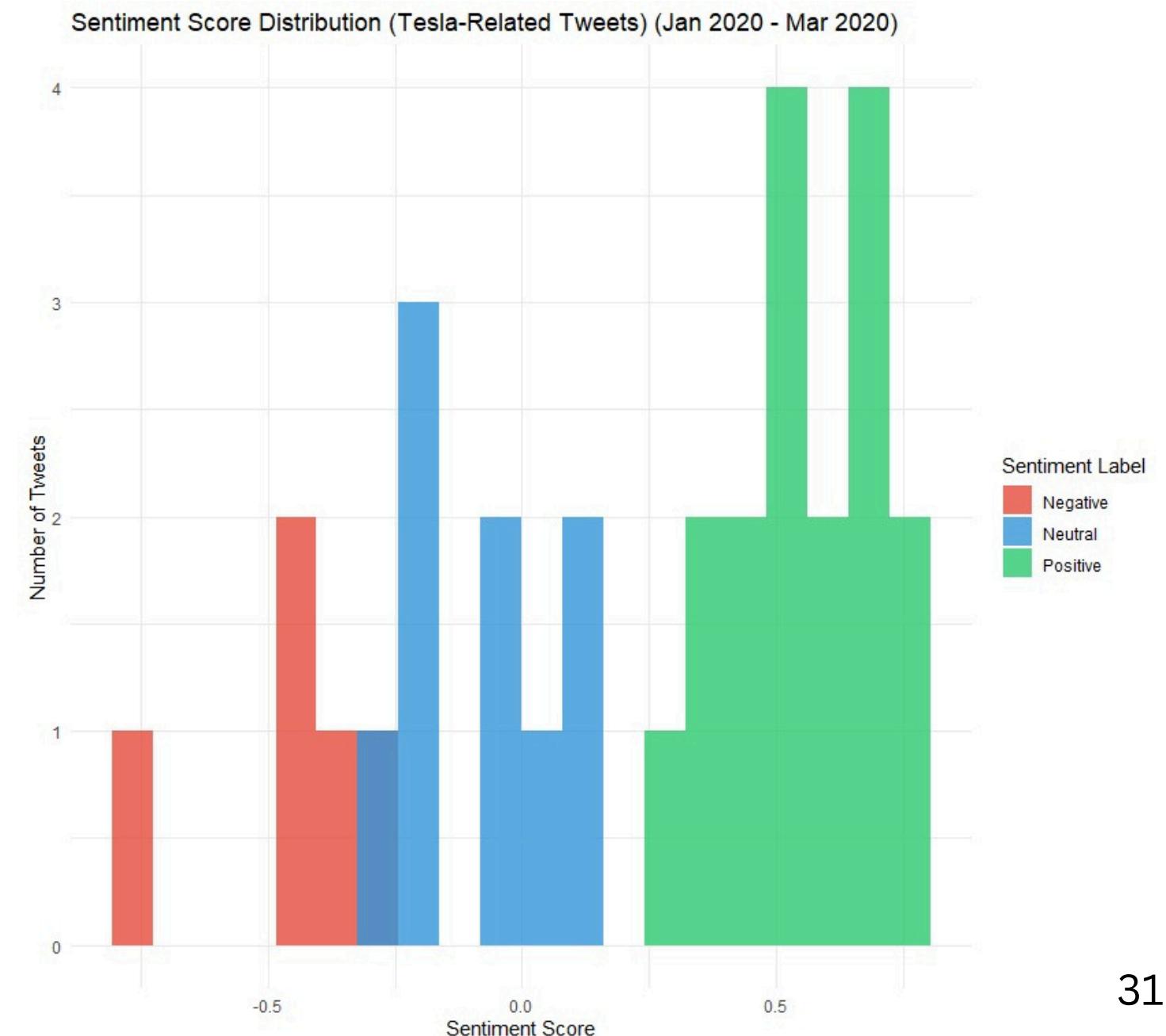
Sentiment – Initial challenges

First Approach: Using the R package tidytext, but this fails to understand context, sarcasm or irony

Second Approach: Using Large Language Model solutions, but this is difficult to use with r studio

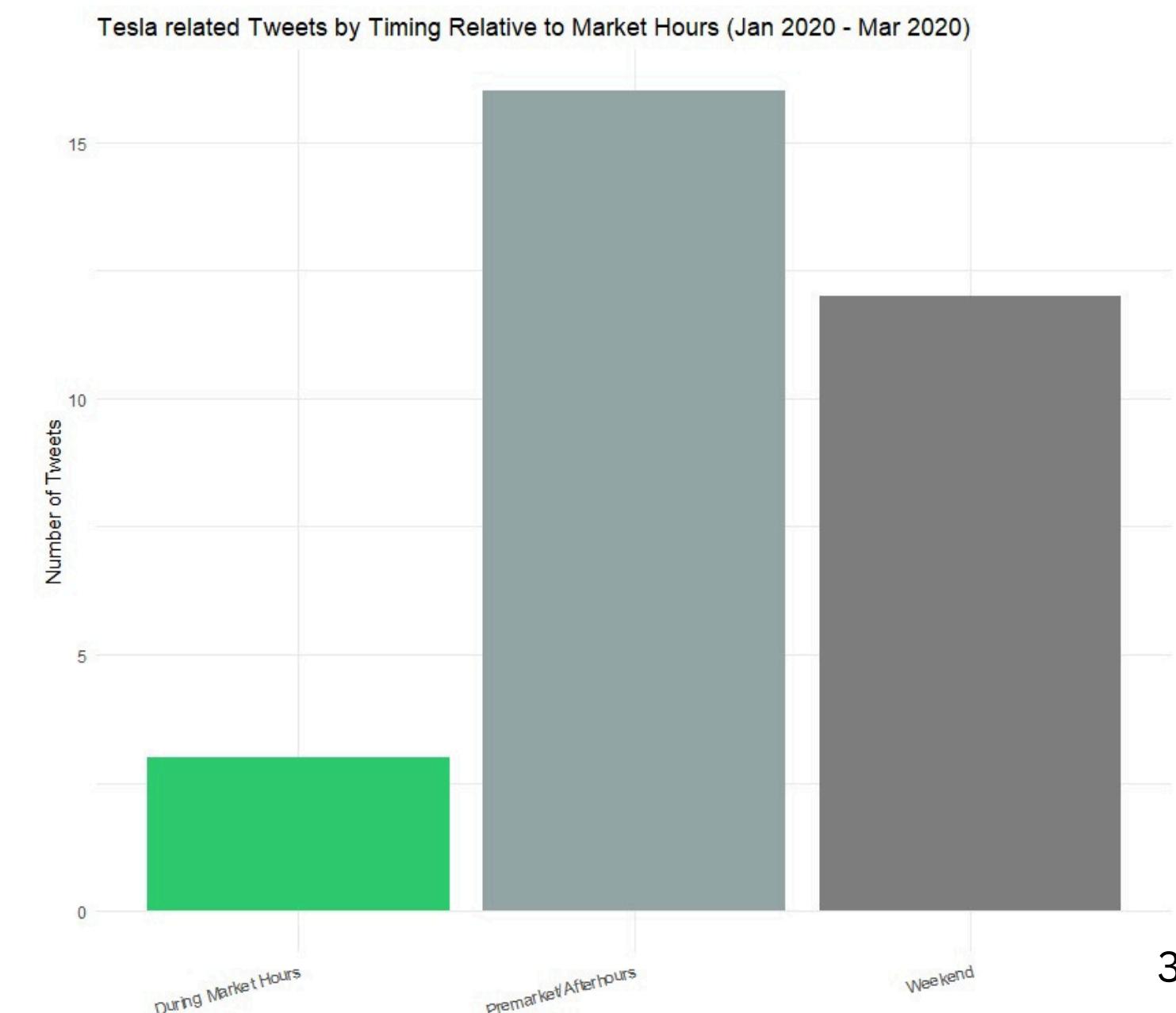
Sentiment- Breakthrough

- Adopted the `sentiment.ai` packages context-sensitive sentiment scoring directly within R
- To support `sentiment.ai`, we had to install Miniconda + Python + TensorFlow



Handling After Hour & Weekend Tweets

- Tweets can occur 24/7, but intraday stock data only exists during U.S. market hours
- 2 Possible approaches:
 - After-hours & weekend tweets : align them with the next market open (09:30)
 - Or just use tweets that occur during the U.S. market hours



Next Steps

- Event-window analysis (-1 to +3 hours)
- Compare reactions across tweets
- Choose analysis approach
- Decide on visualisation strategy
- Organise output data

Feedback & Discussion