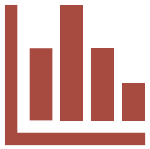
CORK PRESENTATION 1

Maria Benvenuti and Antonio Reche Cazorla

INDEX



Data Cleaning

Steps

Effect on analysis

Effect on data

Database

Type

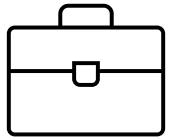
Design

Exploration

Conversations

Definition

Value



Business

Idea

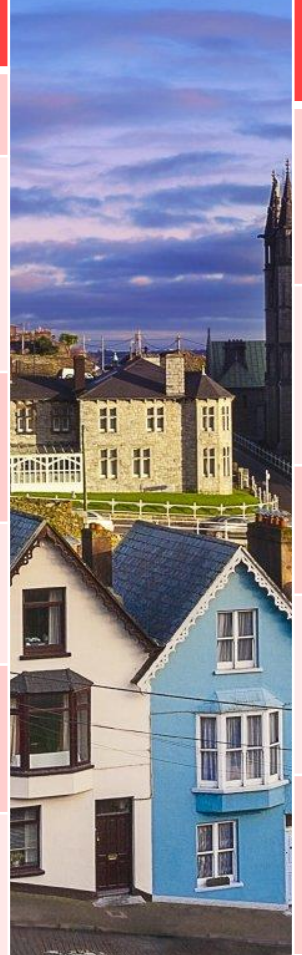
Feasibility

Usefulness

DATA CLEANING

WHICH ATTRIBUTES HAVE WE KEPT

Kept Attributes	Why We Kept Them
text	Core of user input
user.screen_name	Tell us who is interacting with the airline
created_at	Allows temporal analysis
retweet_count	Engagement metric
entities	Provides extra necessary information
id	Unique identifier for each tweet



WHICH ATTRIBUTES HAVE WE NOT KEPT

NOT Kept Attributes	Why We Removed Them
profile_background_color	Purely aesthetic, not used in our analysis
profile_image_url	Purely visual, no value for the analysis
contributors_enabled	Almost always null
geo_enabled	No usable location information
Profile_banner_url	Visual asset, not useful for the analysis

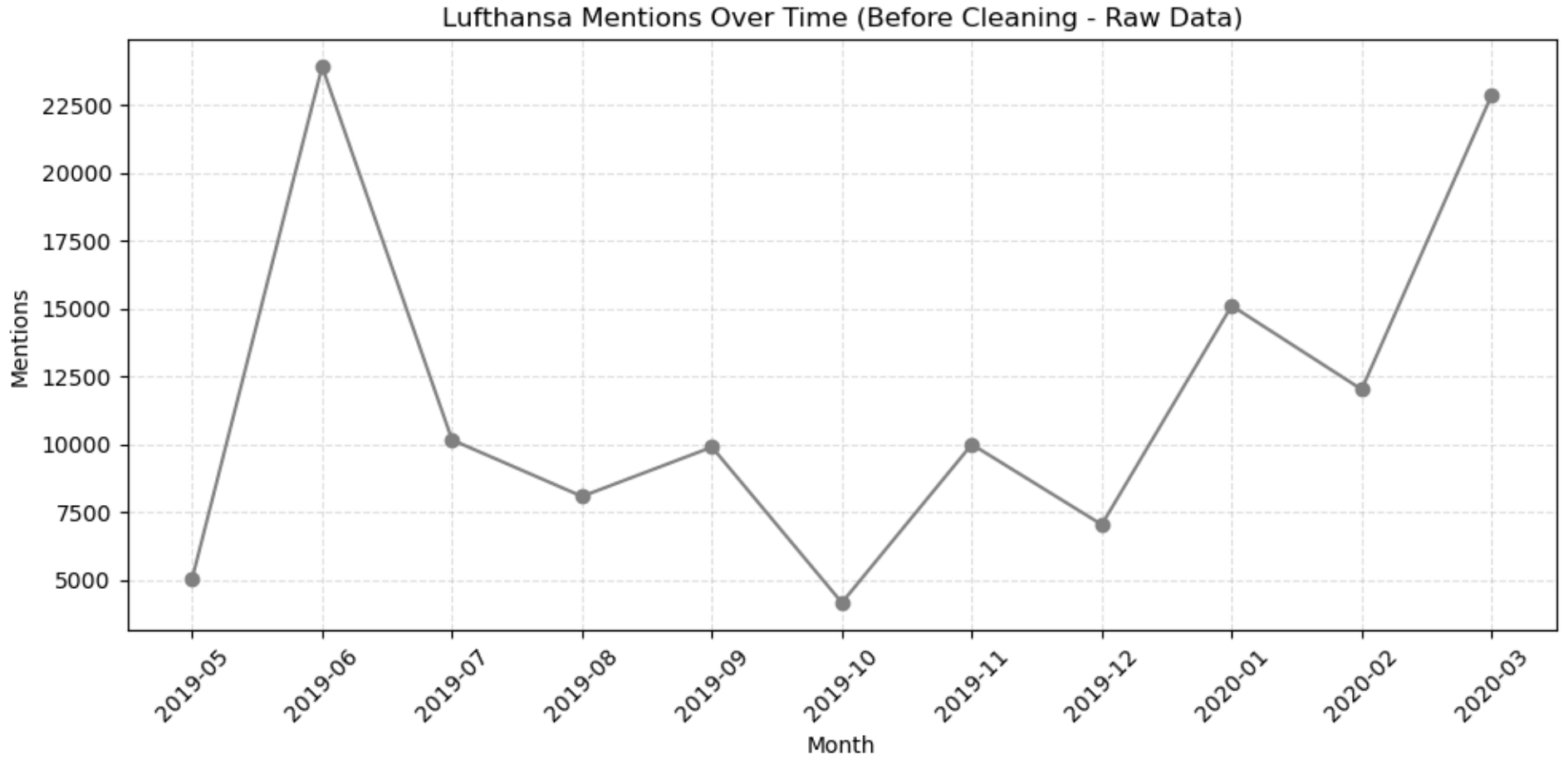
DATA CLEANING: EFFECT ON ANALYSIS

USELESS INFORMATION

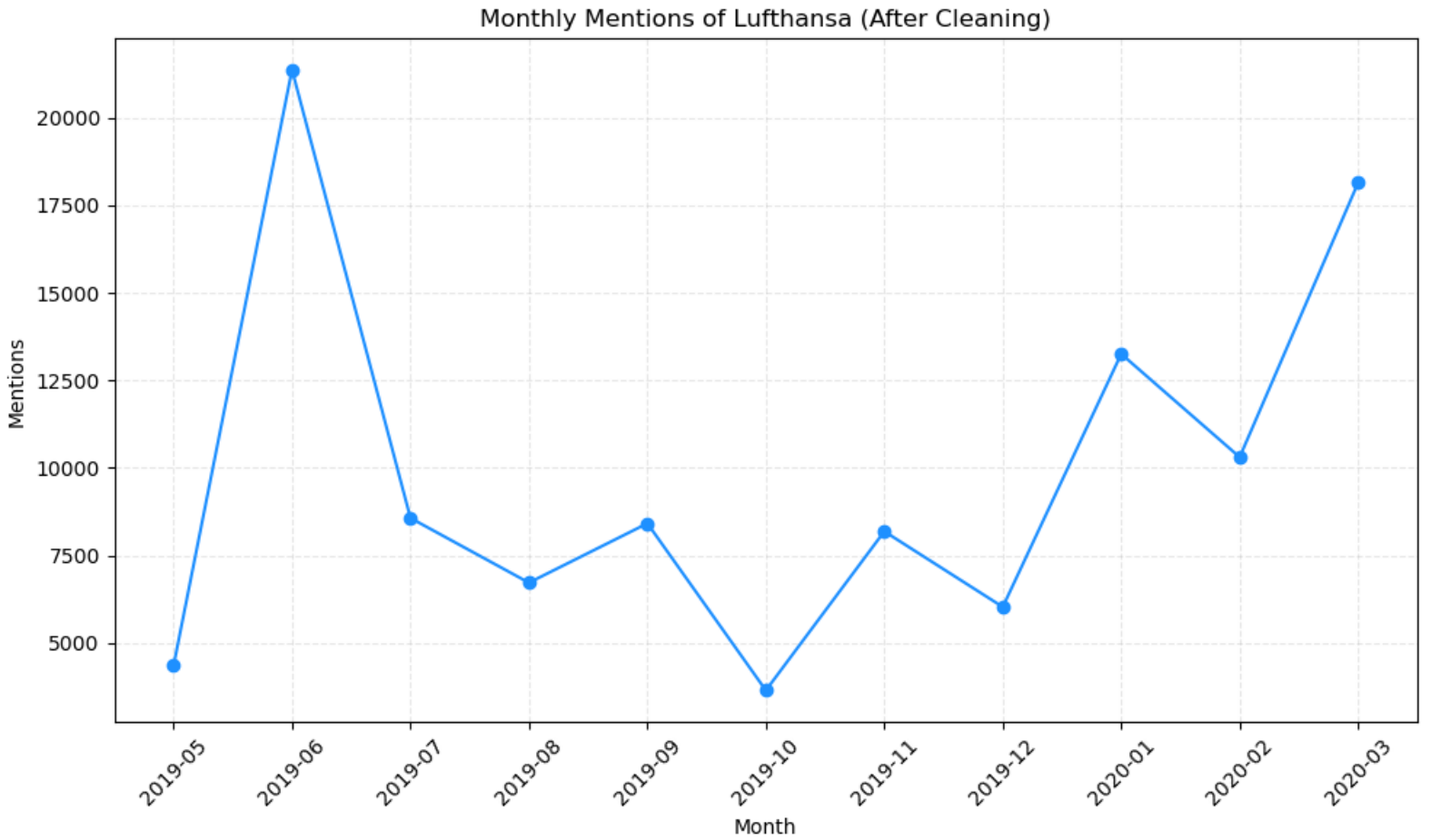
WEIRD/SPAM ACCOUNTS

MISSING INFORMATION

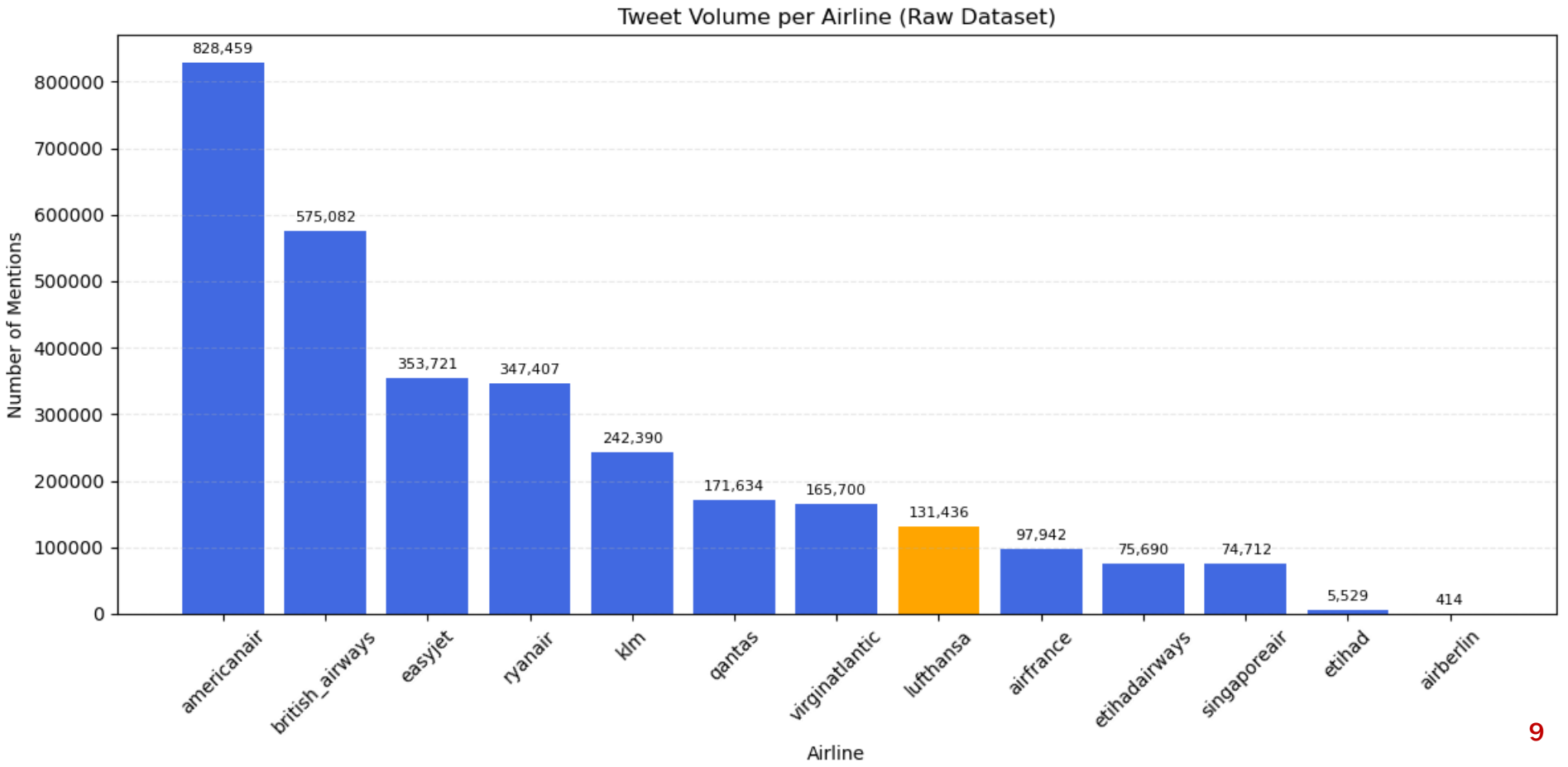
DATA CLEANING: EFFECT ON ANALYSIS



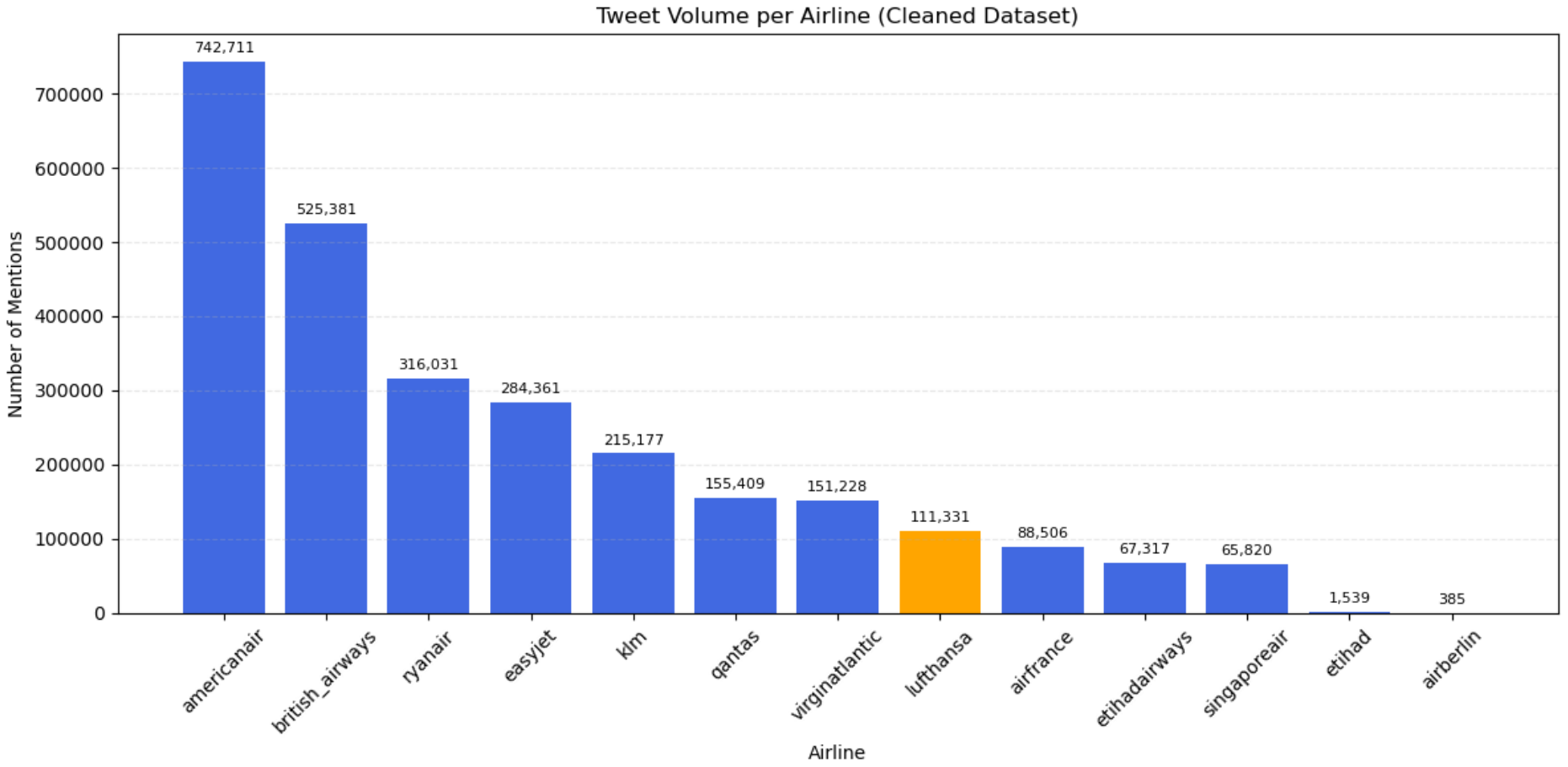
DATA CLEANING: EFFECT ON ANALYSIS



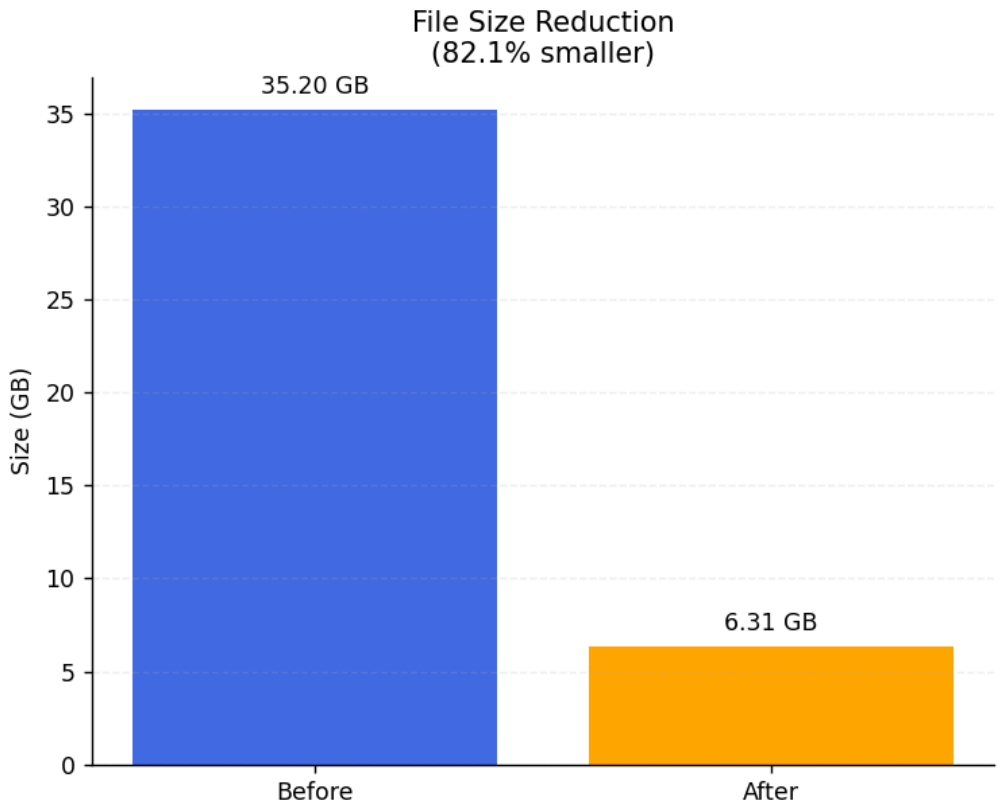
DATA CLEANING: EFFECT ON ANALYSIS

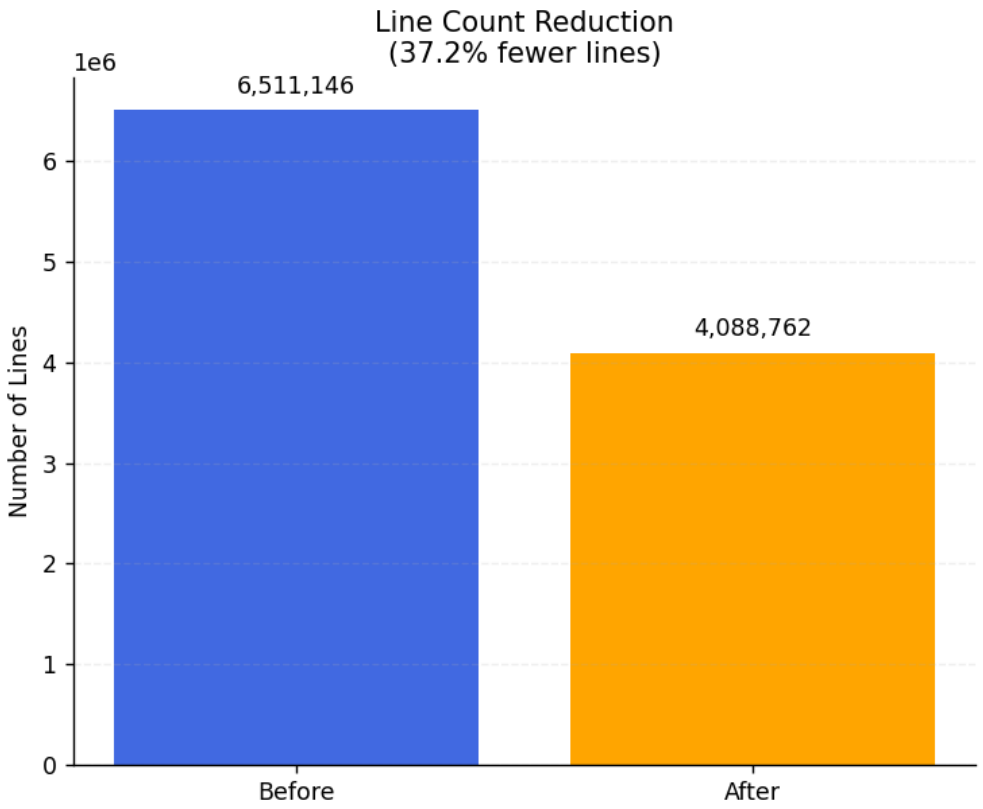


DATA CLEANING: EFFECT ON ANALYSIS



DATA CLEANING: EFFECT ON DATA

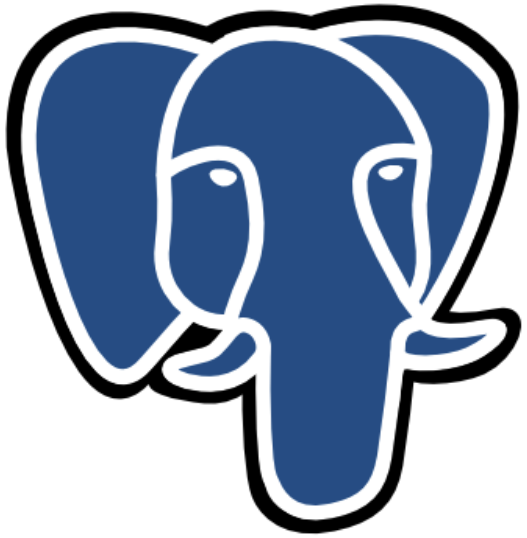




DATABASE

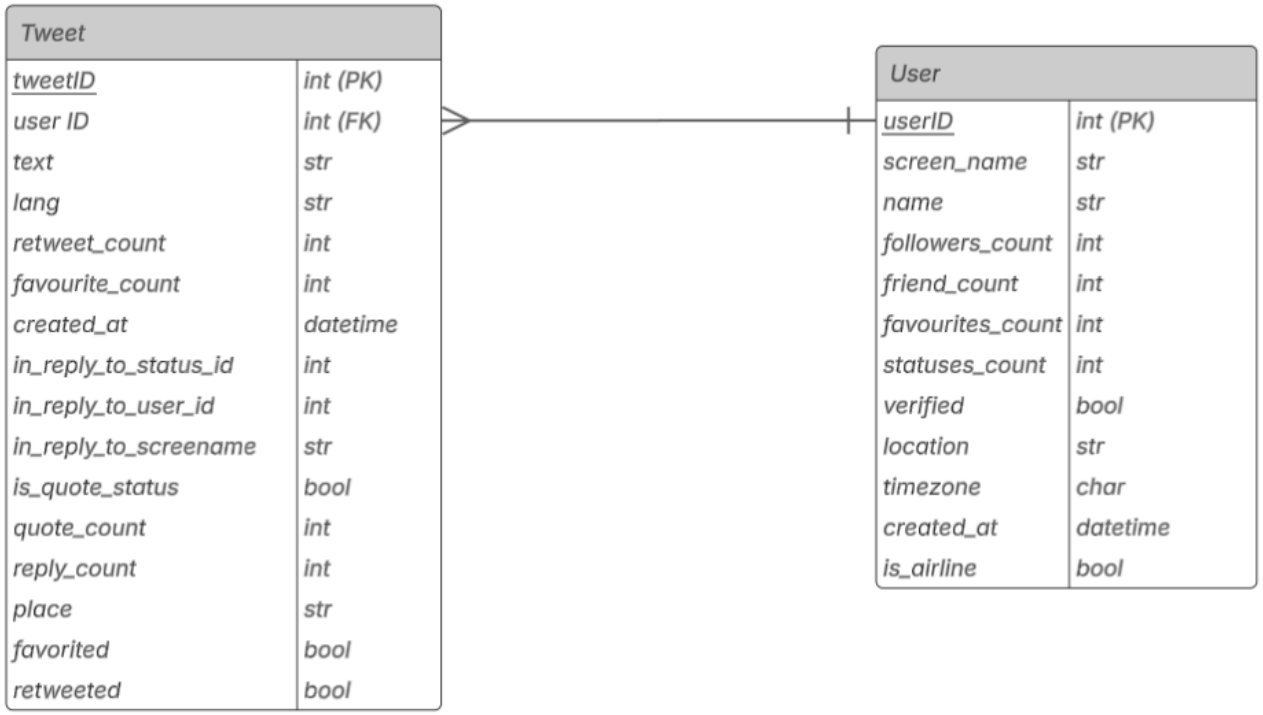
DATABASE: TYPE

- Large volume of data
- Relational database
- Why PostgreSQL?



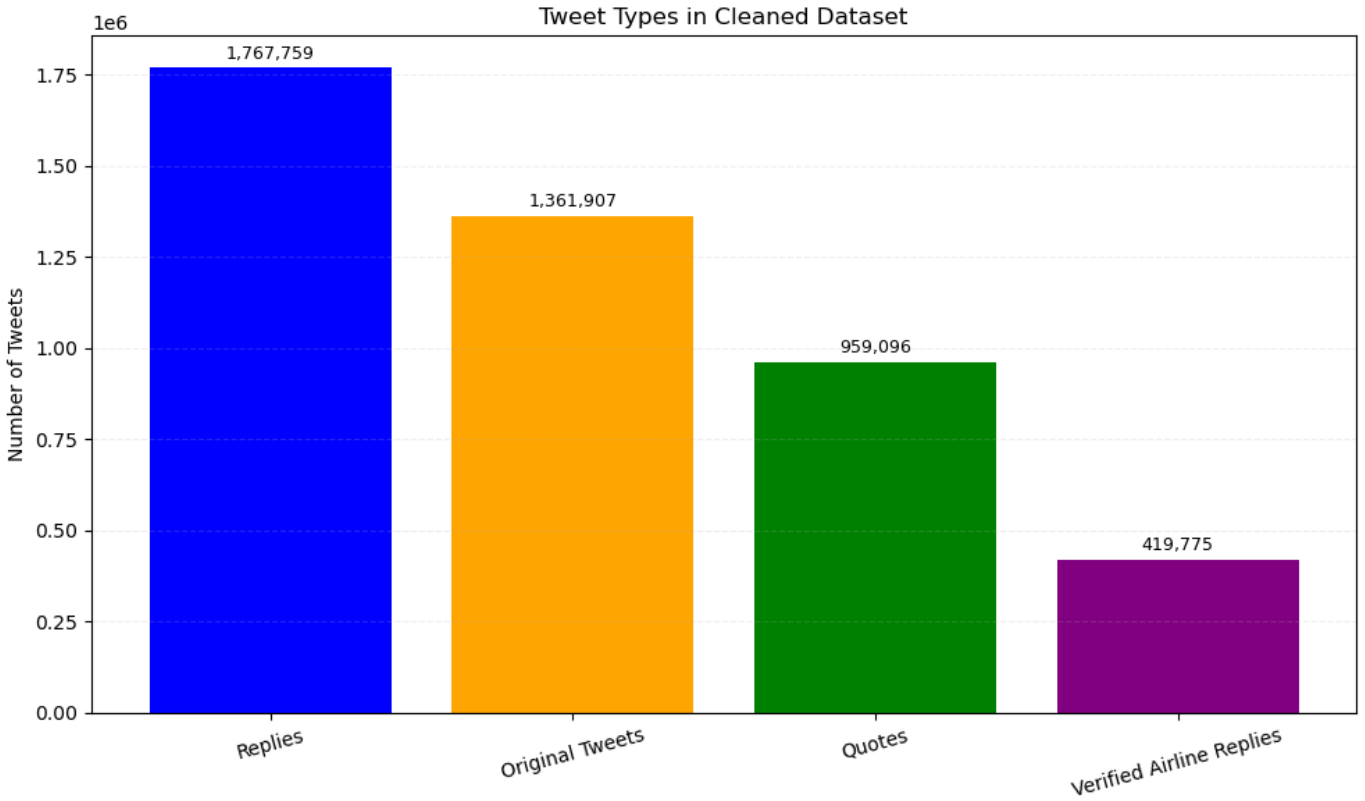
PostgreSQL

DATABASE: DESIGN



- Two table structure schema
- Removed variables not present
- Added extra variable to User table to distinguish between airlines and other users

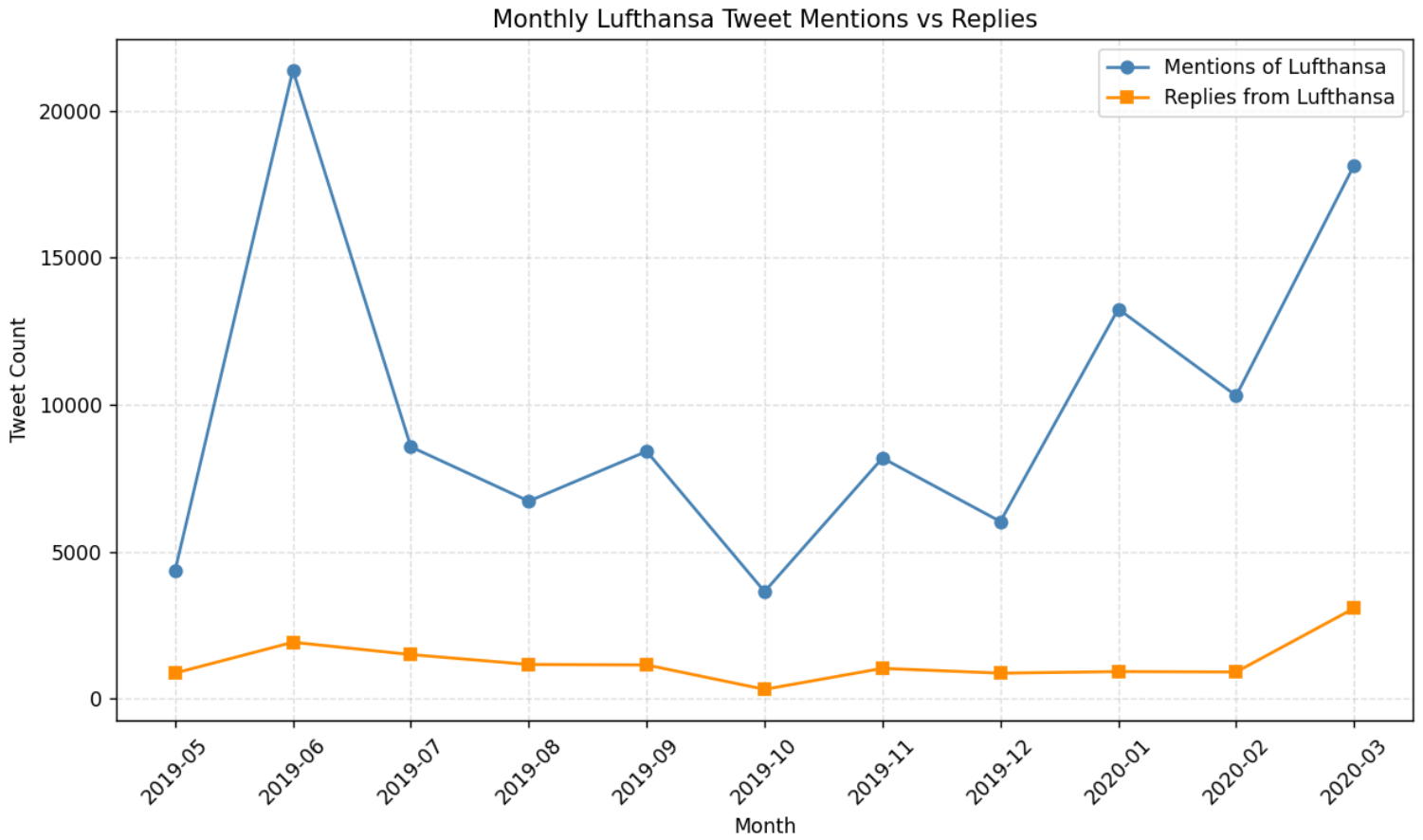
DATABASE: EXPLORATION



Quantity of tweets with:

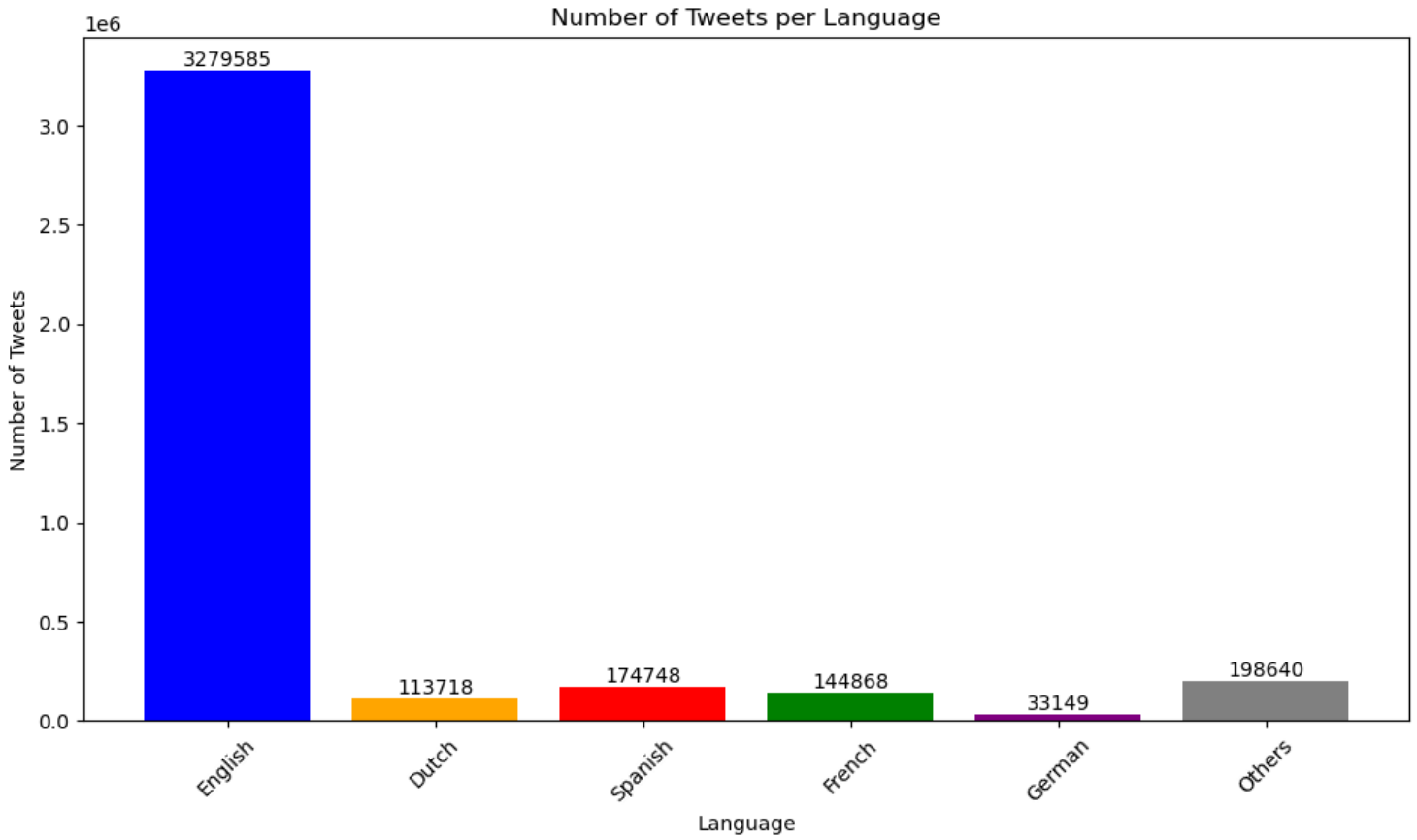
- Replies (Blue)
- Original(Neither reply nor quote) (Yellow)
- Quotes (Green)
- Airline Replies (Purple)

DATABASE: EXPLORATION



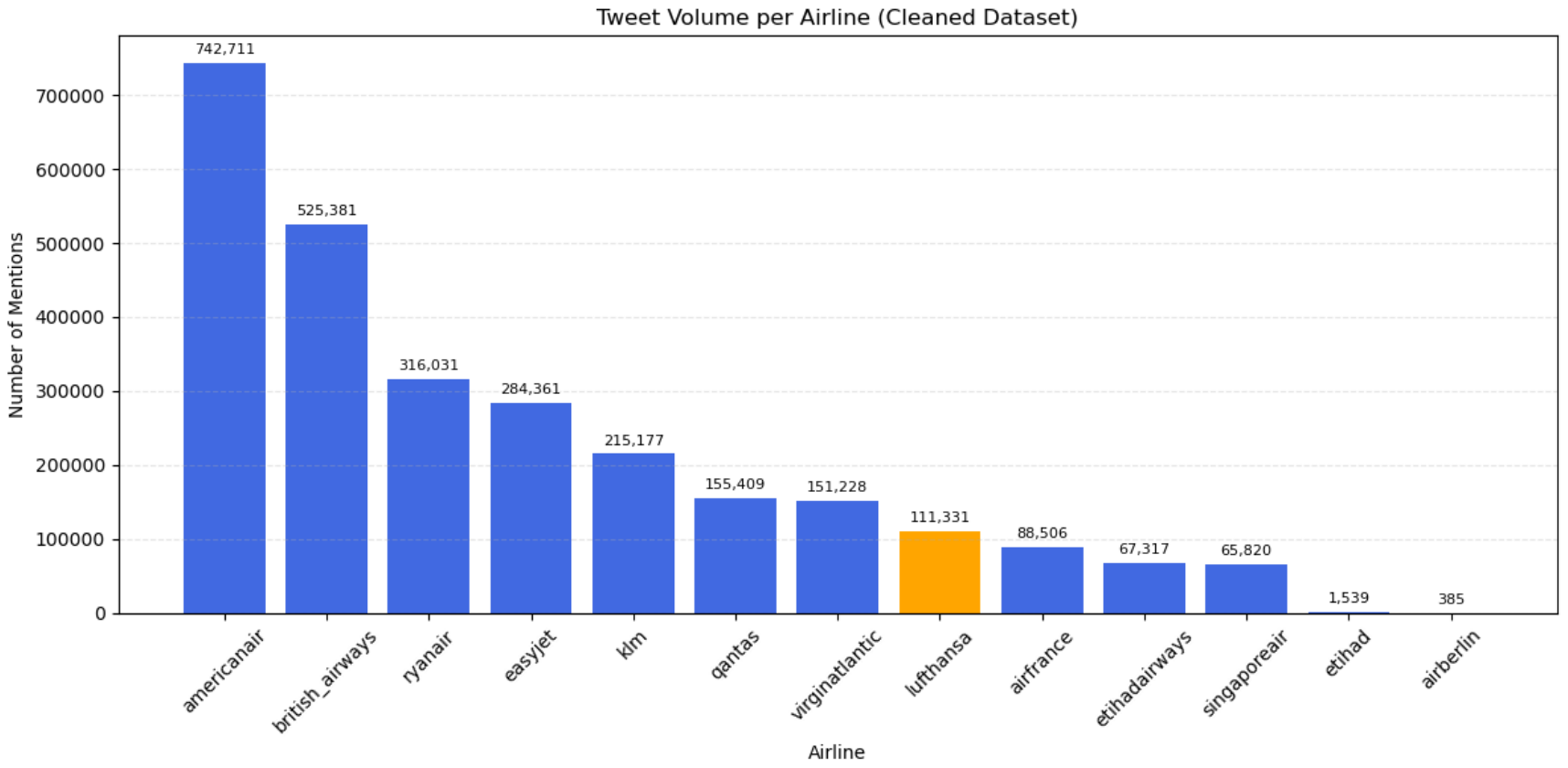
- Lufthansa replied to 13,729 out of 108,959 mentions.
- Overall reply ratio: 12.6%
- 1:7.9 reply to mentions ratio

DATABASE EXPLORATION



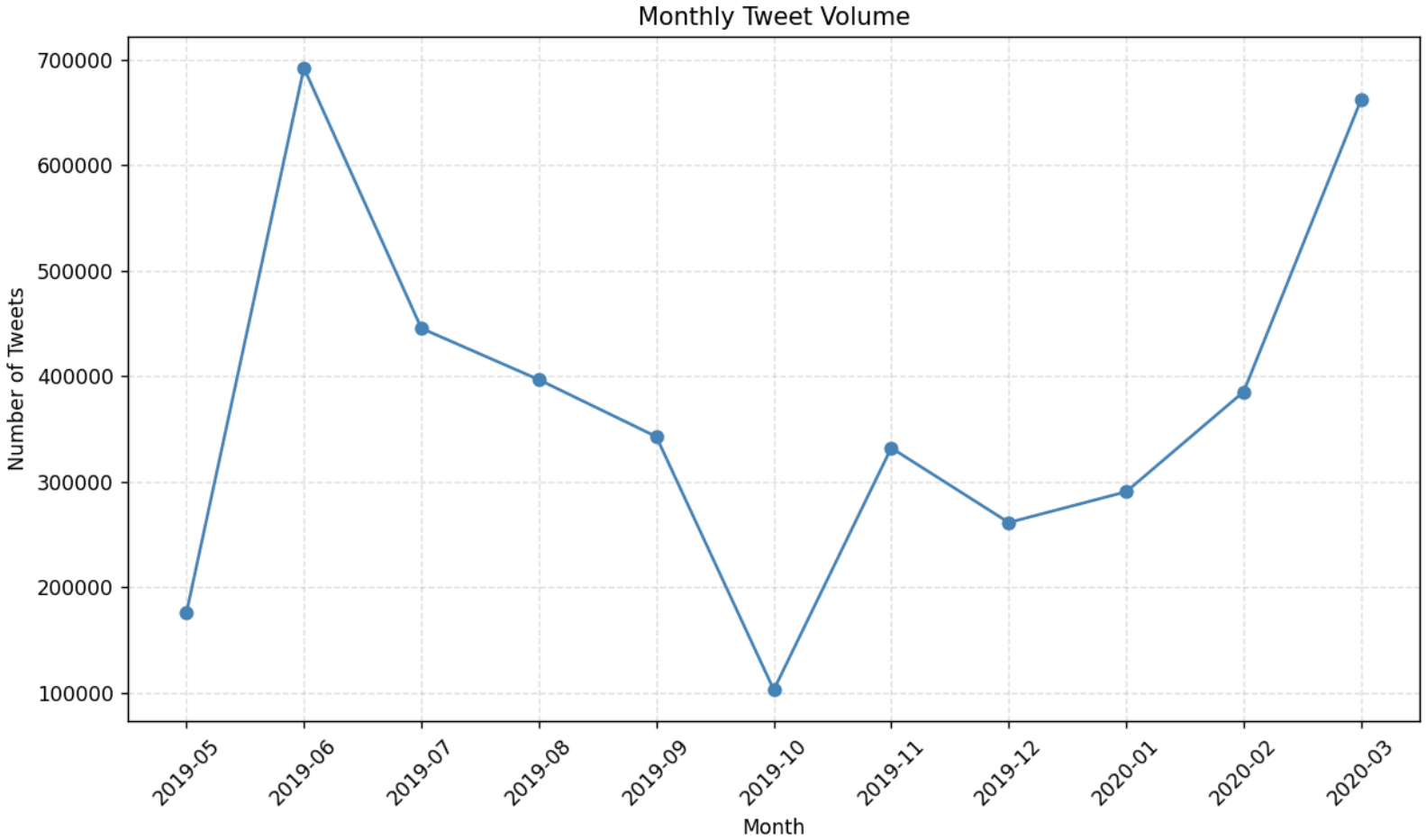
- English dominant, with over 3.2 million tweets
- Global nature of the airline industry and the common use of English in customer service

DATABASE EXPLORATION



- Quantity of tweets received over the given time frame (Lufthansa highlighted in orange)

DATABASE EXPLORATION



- Number of tweets airlines have received over the given timeframe

CONVERSATIONS

CONVERSATIONS: DEFINITION

What is considered as a conversation?

Tweet from a user vs reply from an airline?

EXAMPLE CONVERSATIONS:

Example with reply:

Example with quote:

EXPLANATION CONVERSATIONS

- A situation like this falls under the first example because, to the database this is a reply to a tweet regardless of this being a quote tweet

CONVERSATIONS: VALUE

1,767,759 (43,2%) tweets have received replies

959,096 (23,5%) tweets have been quoted

67% of tweets are part of a “conversation structure”

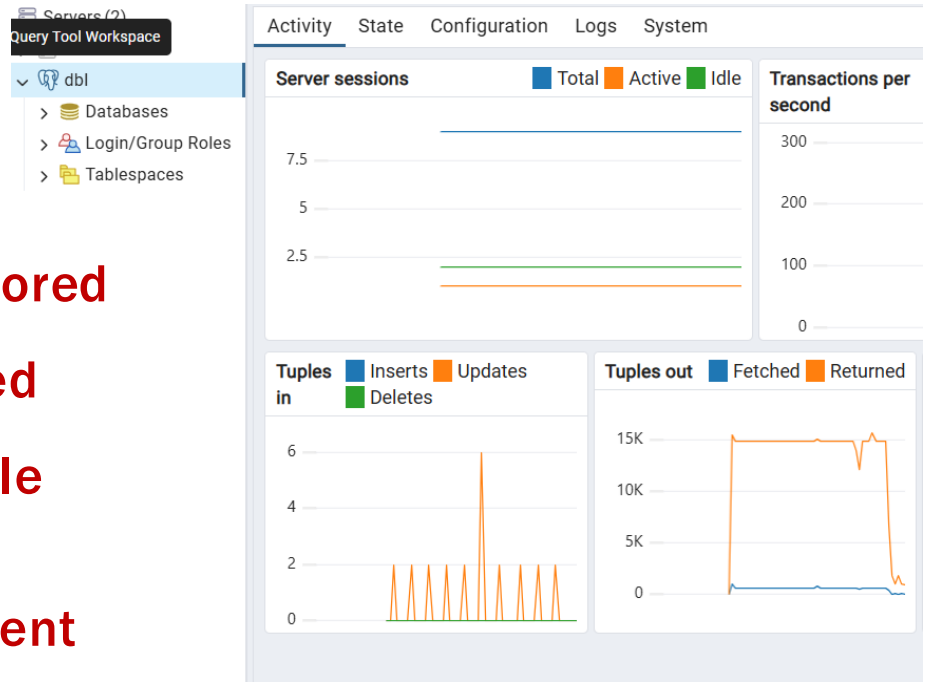
24% of replied tweets are from airlines

BUSINESS

BUSINESS: IDEA

- Improve social media strategy and how it affects the public
- Goal: help Lufthansa shift neutral and critic tweets toward supportive ones, to enhance brand image and customer engagement.
- Methods: Filtering, sentiment analysis, airline comparison and visualization

BUSINESS: FEASIBILITY



- Data cleaned and stored
- Filtering implemented
- Visualizations feasible and reproducible
- Tweet volume sufficient for analysis

```
def is_complaint(text: str) -> bool:···

def is_promotional_spam(text: str) -> bool:···

def extract_relevant_info(tweet: Dict) -> Dict:···

def contains_airline_tag_or_is_quote_or_reply(tweet: Dict) -> bool:···

def plot_tweets_per_language(cleaned_data: List[Dict], save_path: str):···

def is_weird_account(user: Dict, tweet_created_at: str, tweet: Dict) -> Tuple[bool, List[str]]:···

def plot_summary_stats(stats: Dict[str, int]):···

def main():···
```

BUSINESS: USEFULNESS

How is the idea potentially useful?

- Shows how Lufthansa's engagement compares to competitors.
- Highlights weak points in public perception and social media.
- Evaluates effectiveness of Lufthansa's Twitter support team.
- Provides insight for strategic improvements and response tone.
- Links performance to global events and tracks emotional trends over time.

END

EXTRA INFORMATION

Extra list of all the attributes that we have kept

Created_at	Id	Text
Lang	Retweet_count	Favourite_count
In_reply_to_status_id	In_reply_to_user_id	In_reply_to_screen_name
Is_quote_status	Quote_count	Reply_count
Place	Favorited	Retweeted
User	Entities	

EXTRA INFORMATION

Extra list of attributes we have not kept

Id_str	Source	In_reply_to_status_id_str	In_reply_to_user_id_str	Coordinates	Contributors	Extended_tweet
Entities	Filter_level	Timestamp_ms	Extended_tweet	Display_text_range	url	Description
Translator_type	Protected	Listed_count	Utc_offset	Geo_enabled	Contributors_enabled	Is_translator
Profile_background_color	Profile_background_image_url	Profile_background_image_url_https	Profile_background_title	Proflie_link_color	Profile_sidebar_border_color	Profile_sidebar_fill_color
Profile_text_color	Profile-use_background_image	Profile_image_url	Profile_image_url_https	Profile_banner_url	Default_profile	Default_profile_image