Introduction to delivery services



Antonio Salas

DSI 1015

Introduction



Currently, Too Good To Go does not offer a delivery service, limiting our reach to a broader audience and our ability to promote responsible consumption habits.



In 2023, we **sold** a total of **121 million meals**.



While other like **DoorDash** completed over **2161 million deliveries**.



Maybe that's an ambitious goal.

Focus on users of food delivery platforms



DoorDash and UberEats dominate the market with over 80% market share.



Through interactions on the Reddit social platform, we will delve into the profiles of both customers and delivery drivers to better understand their needs and behaviors.

Goal for this project





Analyze the differences in posts from DoorDash and Uber users in specific subreddits to identify key topics and sentiment.

As a future analysis, explore the distinction between drivers and consumers to better understand the experiences of each user type.

Challenges to consider

Reddit as the only information source

Unconsidered competition

Possibility of similar results between companies

2726 post from customers and drivers

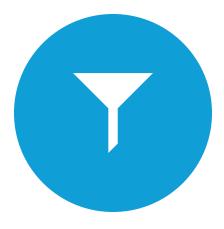


TITLE
SELF_TEXT



NUM_COMMENTS
SCORE
UPVOTE
AUTHOR

MEDIA

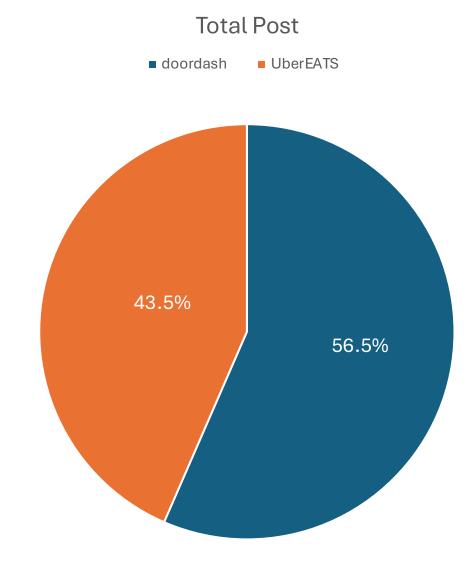


TYPE USER
MAIN SUBJECT

Exploratory Data

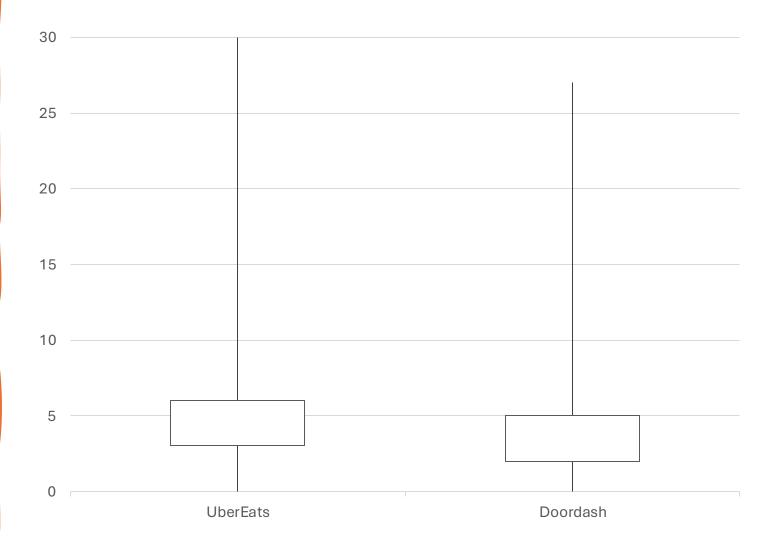
More traffic from r/doordash

 Doordash is the largest food delivery service in the US



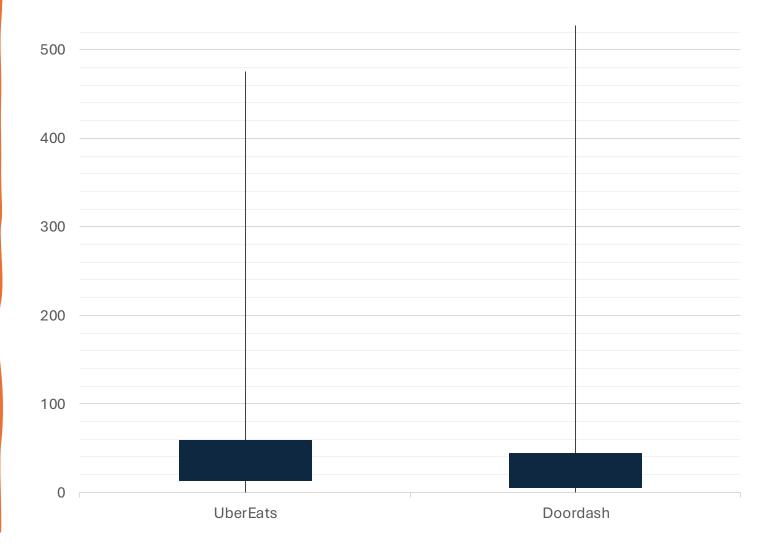
UE* users shared more information

Number of words in Title

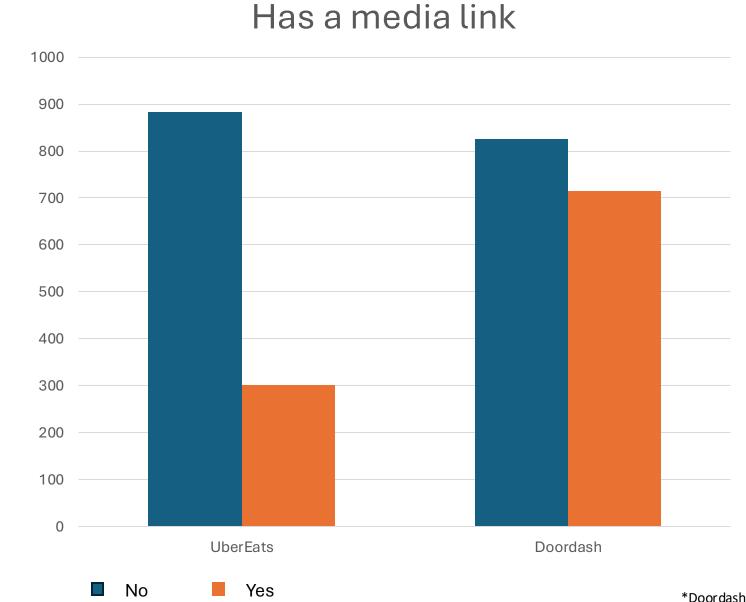


UE* users shared more information

Number of words in Self text

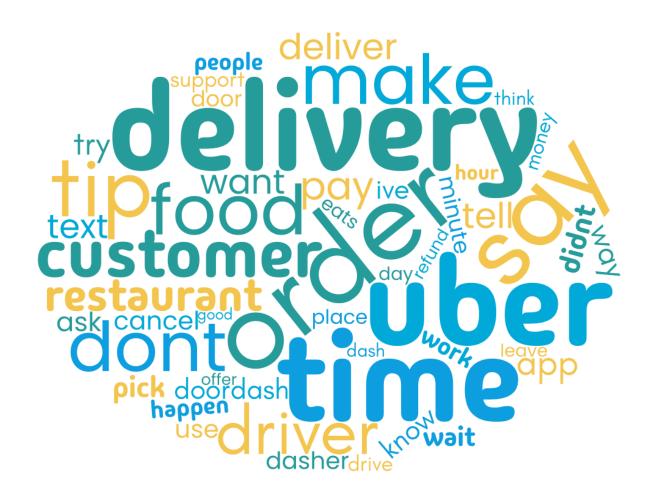


Posts from Dashers* are more visual



And what are they talking about?*

- Tip
- Support
- Refund
- Time
- Customer
- Driver & dasher
- Restaurant



Dasher focus*

- Tip
- Time (hour)
- Driver & dasher
- Support
- Customer
- Restaurant



On Uber we have the same conversation*

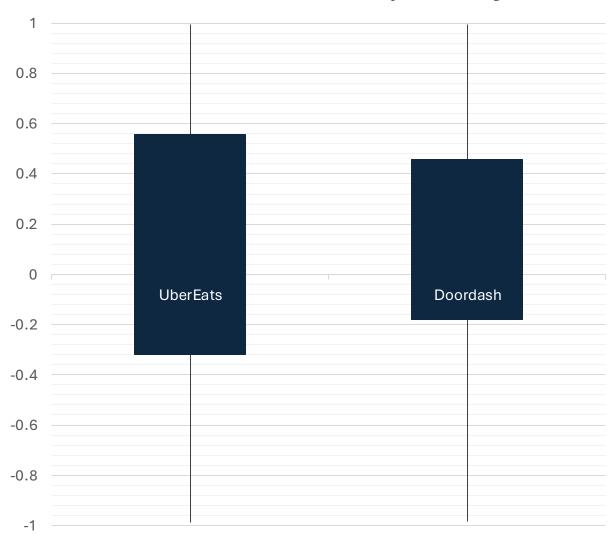
- Drivers
- Customers
- Time (minute, hour)
- Tip
- Restaurant
- Refund
- Support



The message sound neutral

• Although the dispersion of the UE posts is wider, their messages appear neutral, leaning toward positive.

Overall sentiment polarity*



Model & Evaluation

Modeling Approach

- Models Used: Logarithmic Regression & Random Forest
- Data: Combination of title and self_text, word_count combination, has_media
- Predict: subreddit
- Validation procedure or metrics:
 - Accuracy Proportion of correct predictions
 - Precision Proportion of true positives respect all predicted positives
 - Recall true positive rate

Model Comparison

Log. Regression.

Accuracy

81.3%

class	Precision	Recall	F1-Score
UberEATS	85%	69%	76%
Doordash	79%	91%	85%

Random Forest

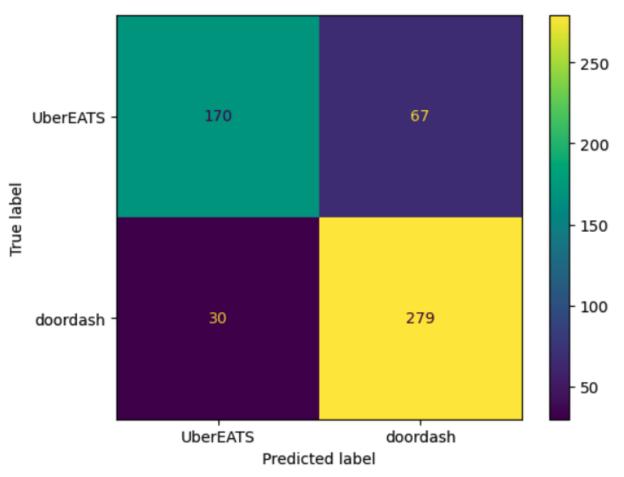
Accuracy

82.2%

class	Precision	Recall	F1-Score
UberEATS	85%	72%	78%
Doordash	81%	90%	85%

For a more visual approach

The Confusion Matrix for Random Forest Model.



Both models identify the same key features as important.

- uber
- doordash
- Dasher
- ubereats
- dash

Findings & next steps

We have a good model to differentiate between app users

- The results are not fully conclusive, as they depend heavily on the presence of app names in the user's post, which are key features for classification.
- We recommend expanding this analysis by categorizing posts based on their main subject, to better understand the underlying themes and improve classification accuracy.

Change the classification goal

- At the beginning of this project, our team focused on classifying posts based on whether the author was a driver or a customer.
- This approach will provide valuable insights into different user groups and help us prepare for future market entry.

Other techniques

- Find more advanced techniques for text analysis
- Use an ensemble learning (combine different model to create a more robust one)

Thank you