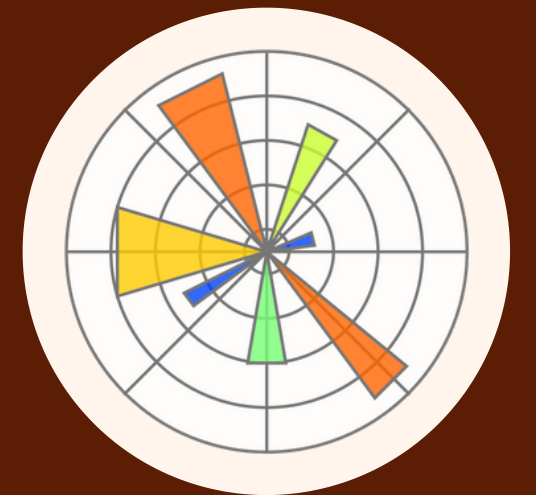USER & BEER TYPE

BEER REVIEWS

samantha rood
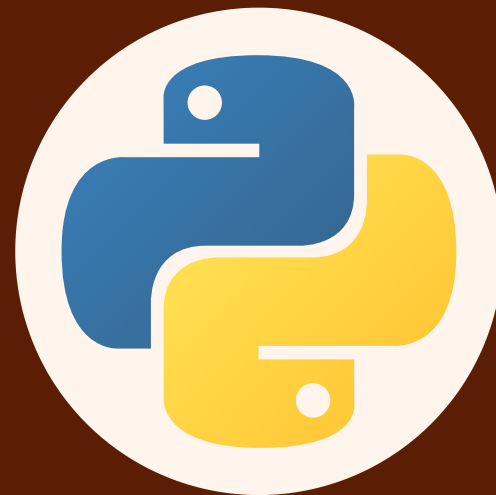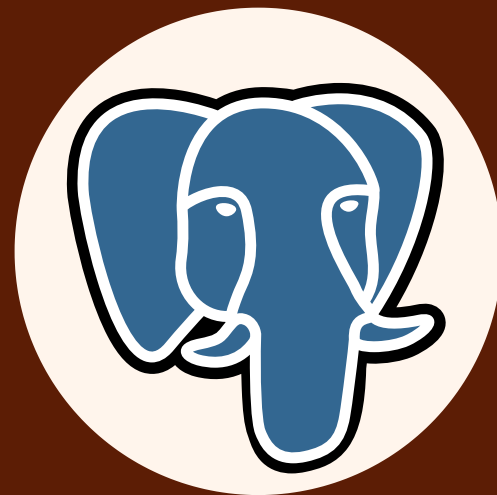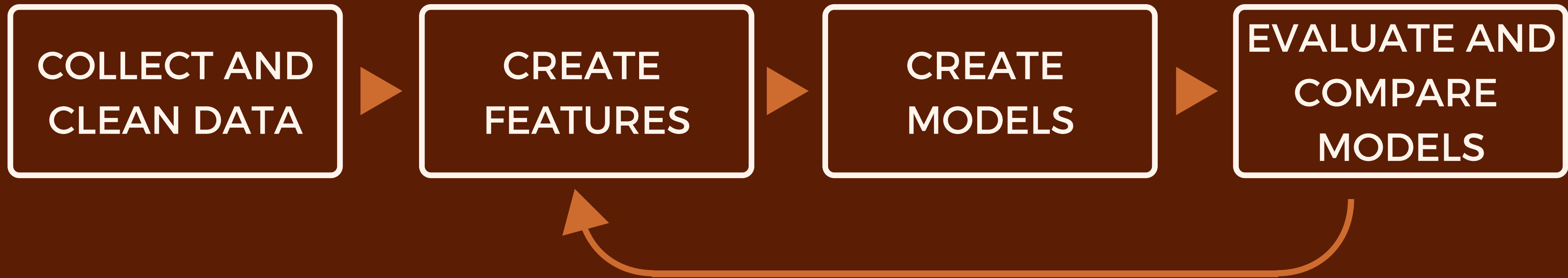
Intro

Can we predict if a user will enjoy a beer type based on their reviews of beer types?

Methodology

| COLLECT AND CLEAN DATA | ▶ | CREATE FEATURES | ▶ | CREATE MODELS | ▶ | EVALUATE AND COMPARE MODELS |

Features

○ **User** — averages of reviews by user, counts of user reviews

○ **Beer Type** — averages of reviews of beer types, counts of beer type reviews

○ **User-Type** — averages and counts of reviews and rating by user of each beer type

Results

all models

ROC curve

logistic regression: area = 0.95
naive bayes: area = 0.86
knn: area = 0.95
decision tree: area = 1.0
random forest: area = 1.0
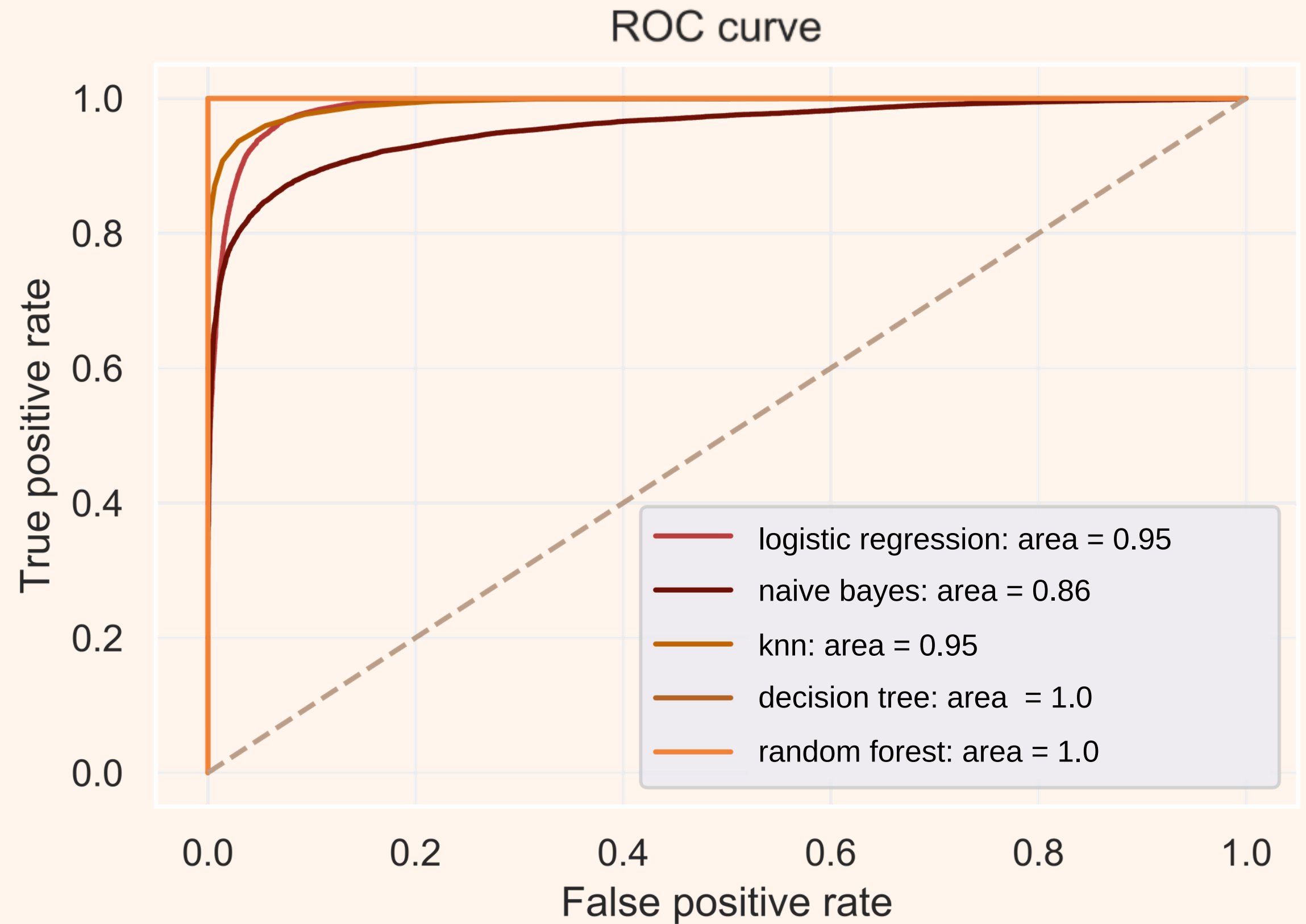
Results

logistic regression scores

Accuracy	0.95
ROC AUC	0.99
F 1	0.95
Precision	0.94
Recall	0.96

Results

feature importance

Relative Feature Importance for Logistic Regression

Feature	
user_type_overall_rating_diff	100
user_avg_overall_rating_diff	~54
type_avg_rating	~51
user_avg_overall_rating_of_type	~50
type_avg_taste_rating	~27
user_review_type_ratio	~26
user_avg_taste_rating_of_type	~16
taste_rating_diff	~10
user_overall_avg_rating	~4
user_avg_aroma_rating_of_type	~3
user_avg_palate_rating_of_type	~3
user_review_number_of_types	~2
user_avg_appearance_rating_of_type	~1
user_type_review_count	0

Future Work

- **Feature selection**

- **Further research**

- **Feature engineering**

Thank you

Appendix

Feature Description

DESCRIPTION

1. Total number of reviews the user has done
2. Number of reviews for each type done by the user
3. Ratio of reviews of type compared to total reviews done by the user
4. Number of different types of beers user has reviewed
5. Average overall rating for each unique beer type
6. Average aroma rating for each beer type
7. Average appearance rating for each beer type
8. Average palate rating for each beer type
9. Average taste rating for each beer type
10. Difference between the users average overall rating and the users average overall rating for a type
11. Difference between the users average overall rating of a beer type and that types overall average rating
12. Average overall rating for each type
13. Total number of reviews for each beer type
14. Average taste rating for each type
15. Difference between the users average taste rating of a beer type and that types taste average rating
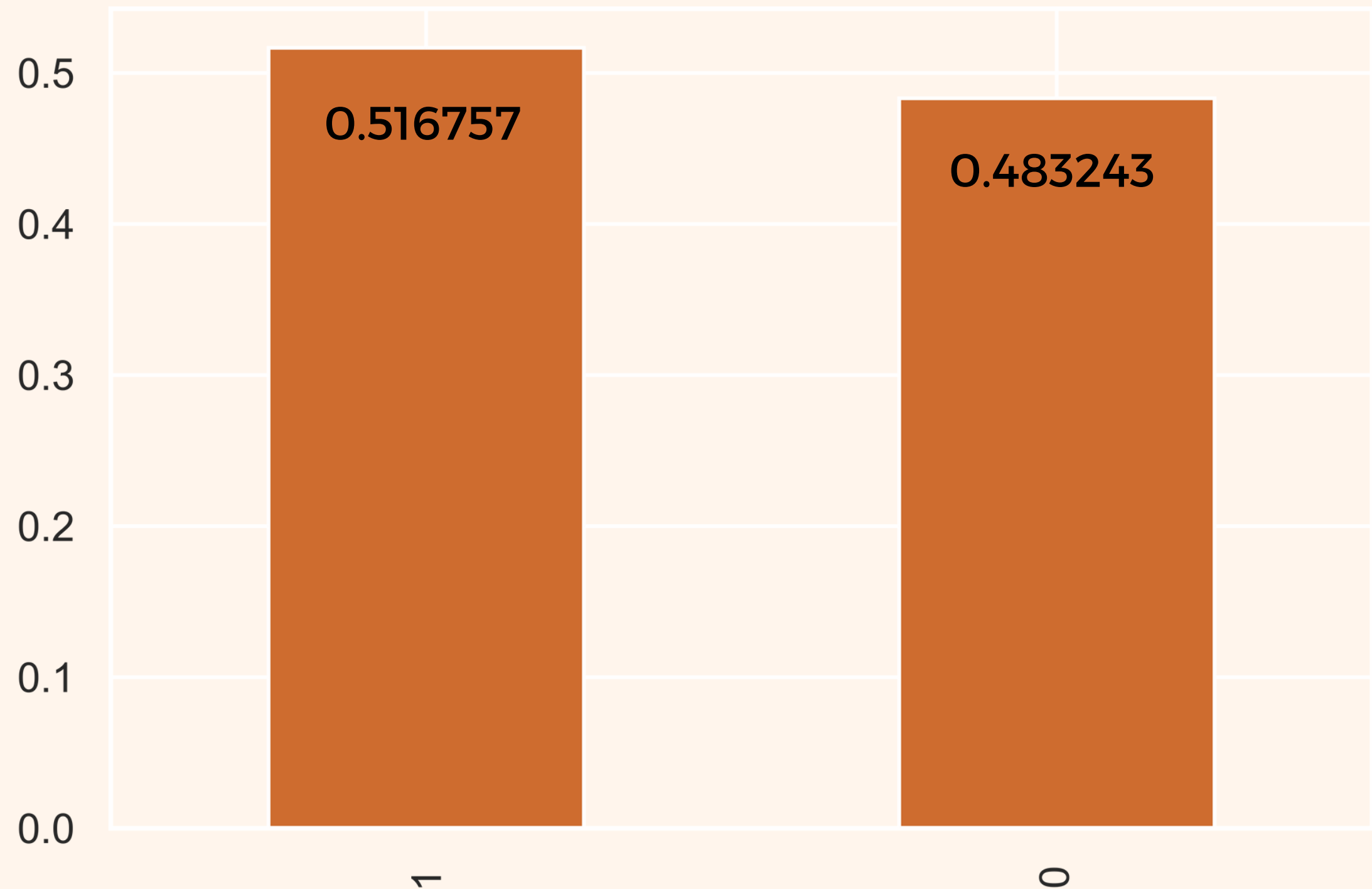
LABEL

1. user_total_review_count
2. user_type_review_count
3. user_review_type_ratio
4. user_review_number_of_types
5. user_avg_overall_rating_of_type
6. user_avg_aroma_rating_of_type
7. user_avg_appearance_rating_of_type
8. user_avg_palate_rating_of_type
9. user_avg_taste_rating_of_type
10. user_avg_overall_rating_diff
11. user_type_overall_rating_diff
12. type_avg_rating
13. type_total_review_count
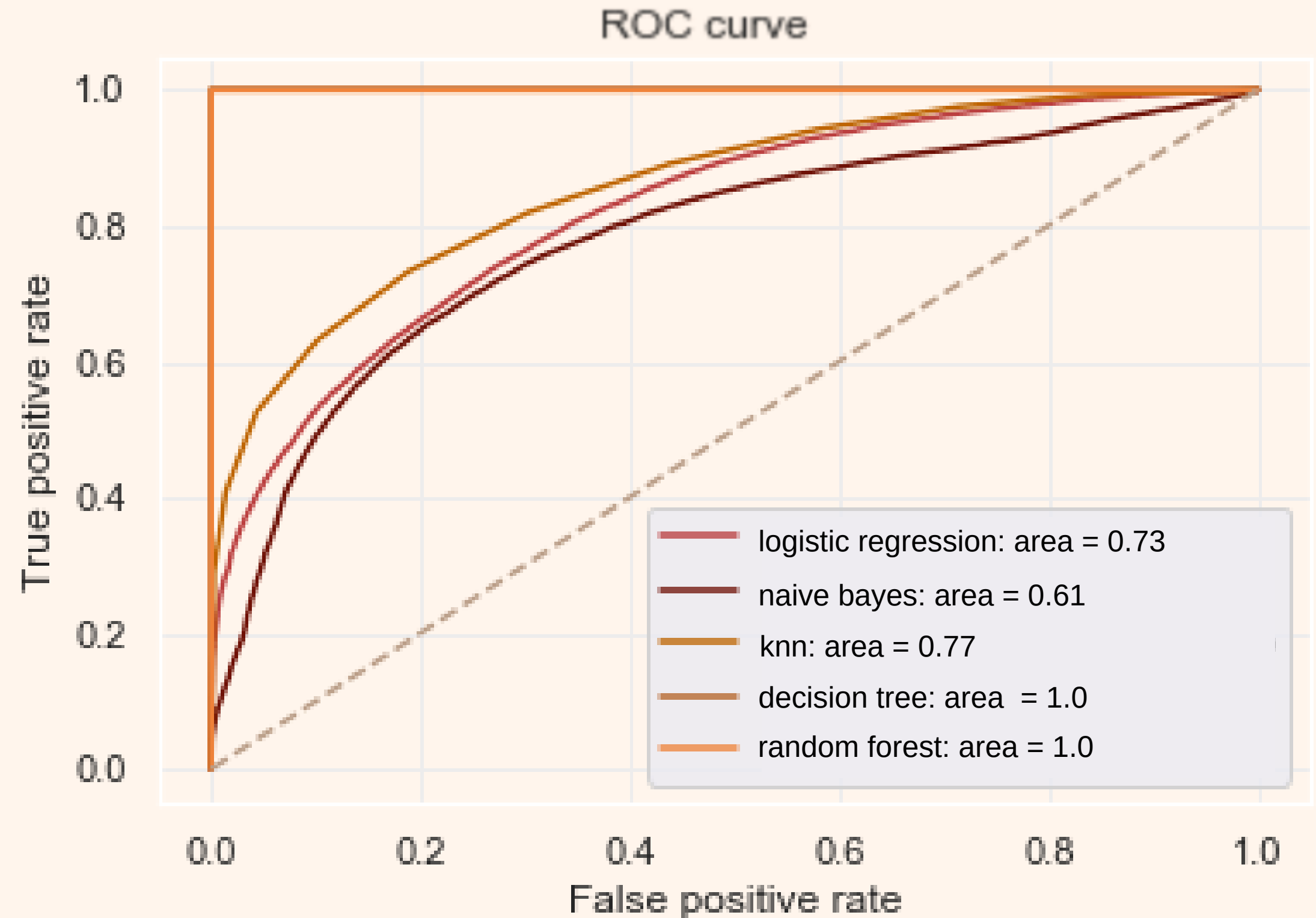14. type_avg_taste_rating
15. taste_rating_diff

ROC Curve For All Models

From the first set of features

ROC curve

True positive rate

False positive rate

logistic regression: area = 0.73
naive bayes: area = 0.61
knn: area = 0.77
decision tree: area = 1.0
random forest: area = 1.0

Logistic Regression Feature Importance

From the first set of features

Feature	
user_avg_overall_rating_of_type	
type_avg_rating	
user_avg_taste_rating_of_type	
user_avg_palate_rating_of_type	
user_overall_avg_rating	
user_avg_aroma_rating_of_type	
user_review_number_of_types	
user_review_type_ratio	
user_avg_appearance_rating_of_type	
user_type_review_count	
user_total_reviews	
type_total_review_count	

Relative Feature Importance for logistic regression
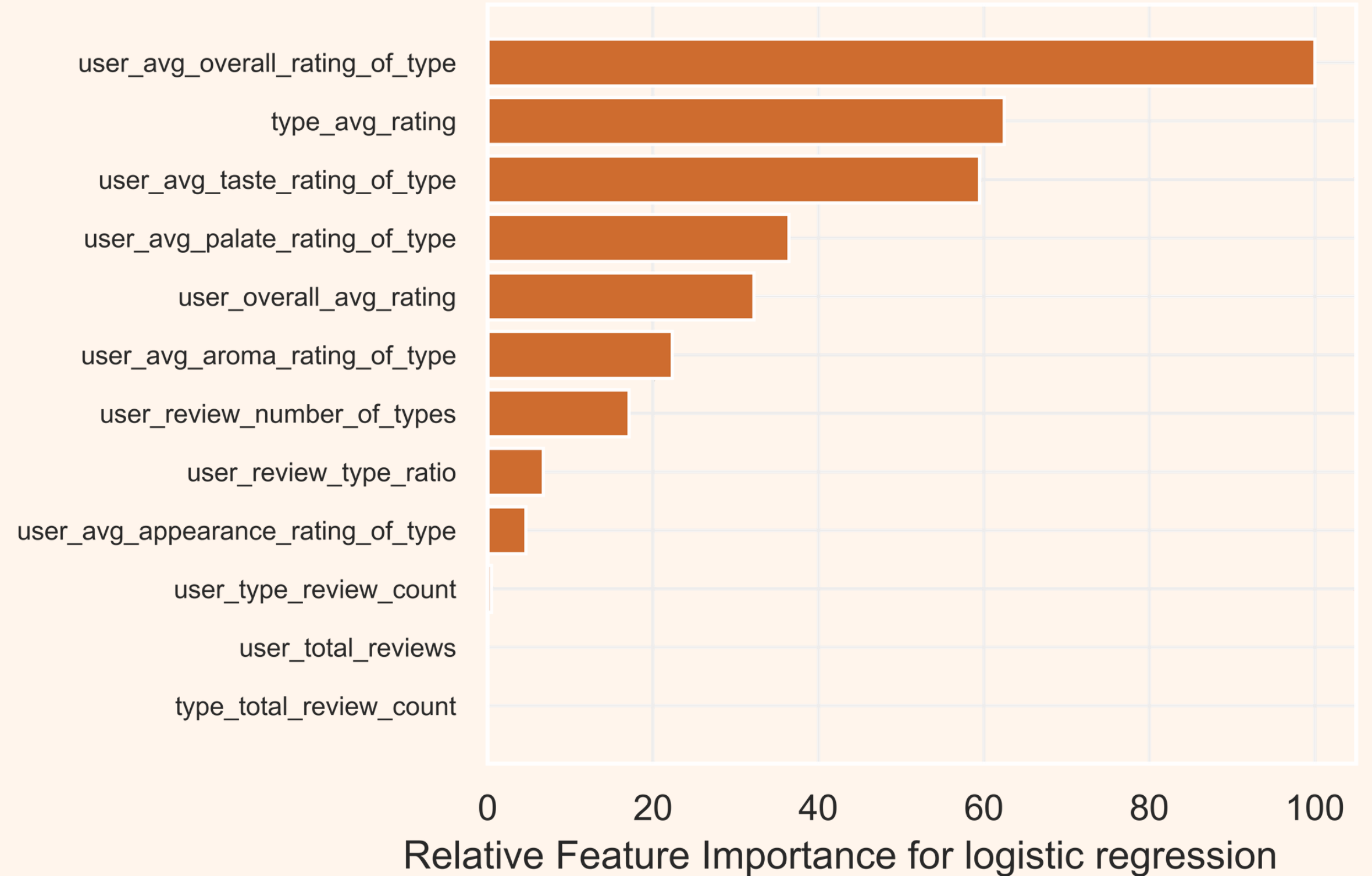
Logistic Regression Confusion Matrix

From the final set of features

Confusion Matrix of Logistic Regression
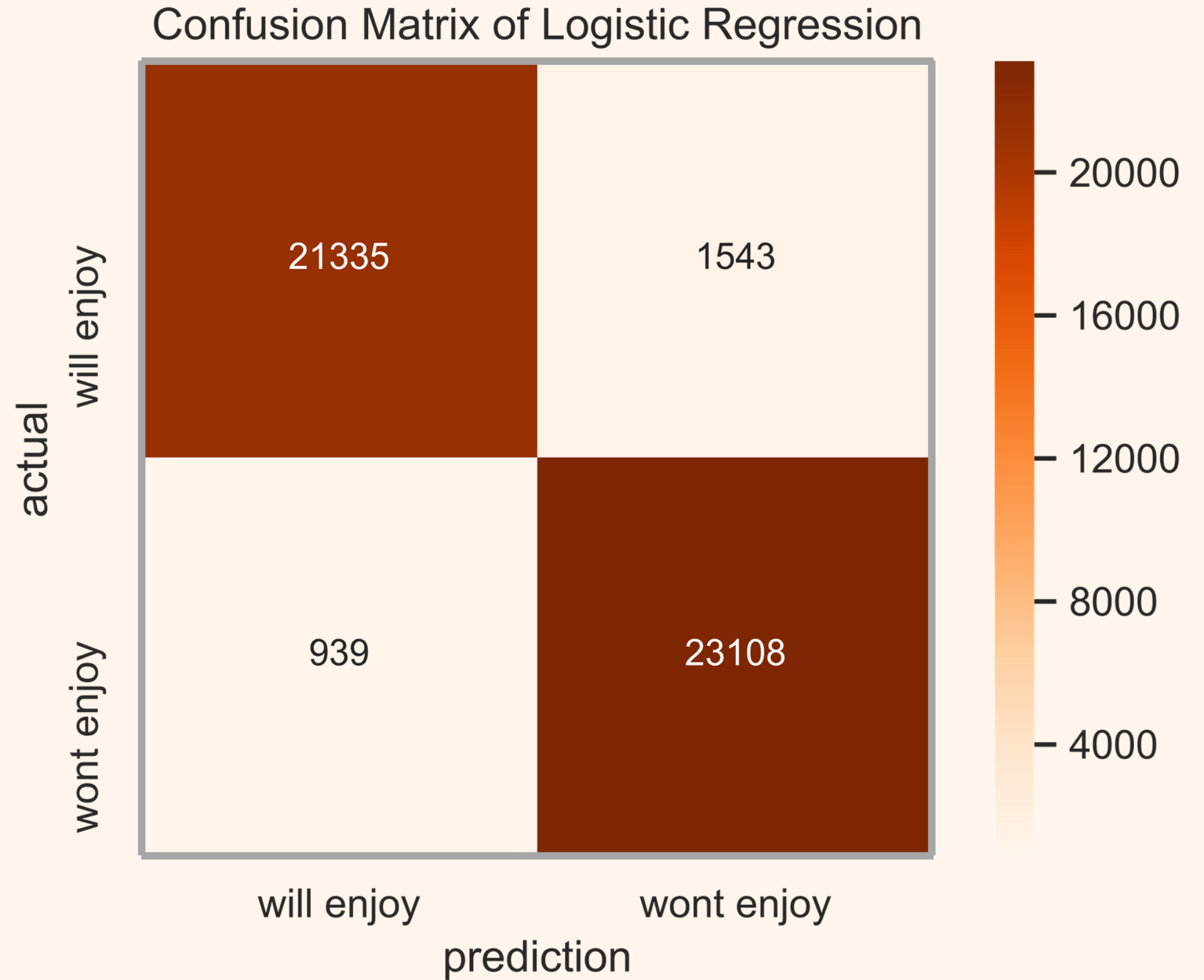
Logistic Regression ROC Curve

From the final set of features

ROC Curve for Will Enjoy on Logistic Regression
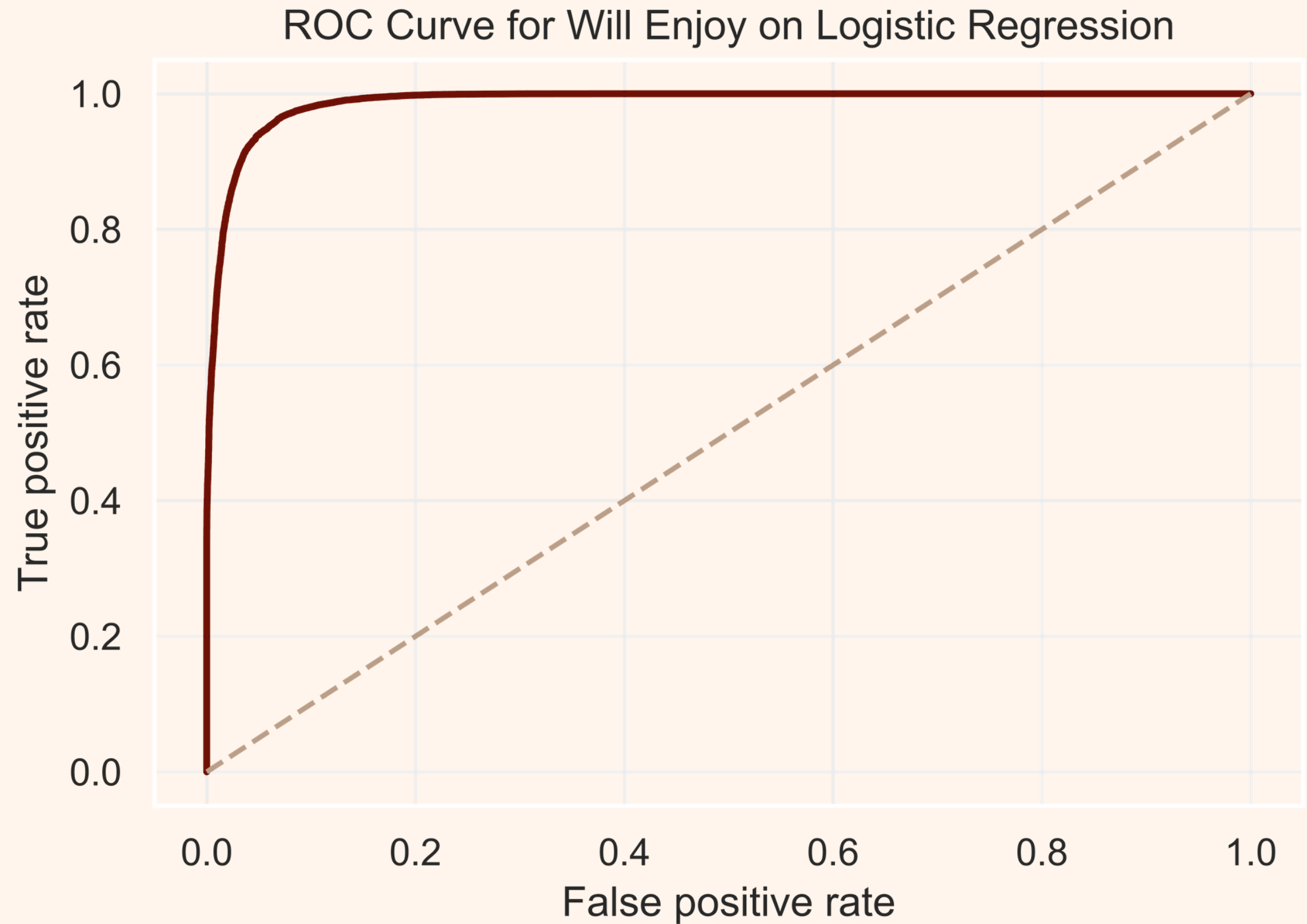
Logistic Regression Recall and Precision

From the final set of features

Precision and Recall Curves for Logistic Regression