



ONLINE SHOPPER'S INTENTION

LAUREN CUNNINGHAM, AND DIPTA
ROY

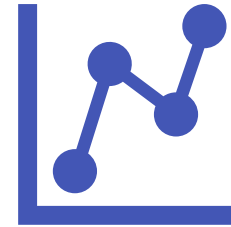
OVERVIEW



What should businesses focus on, to get the highest customer engagement?



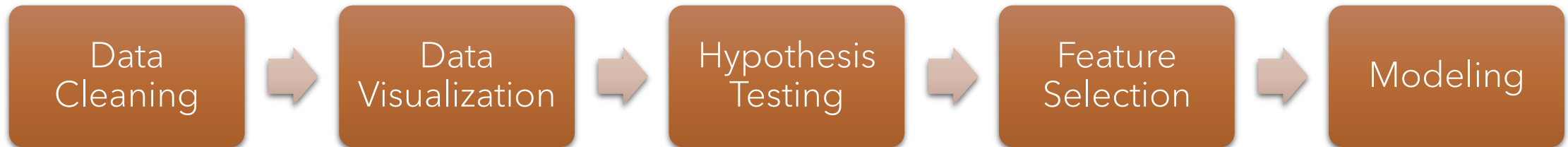
What features are the most indicative of a purchase?



Can we predict if a visitor will make a purchase?

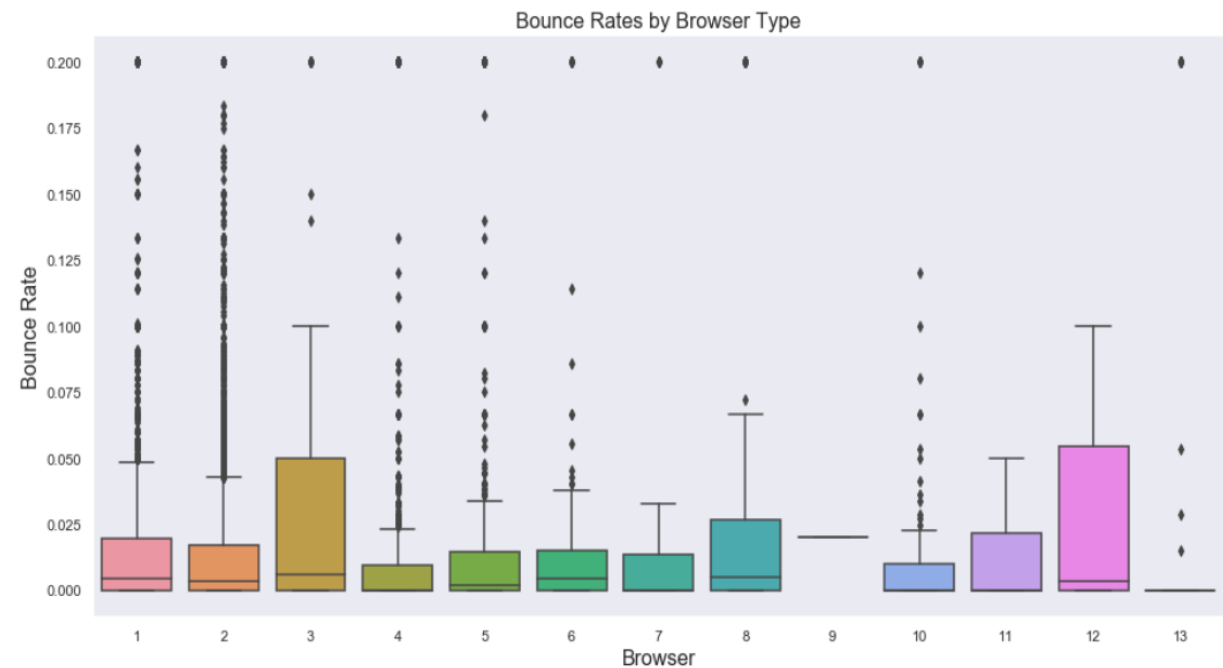
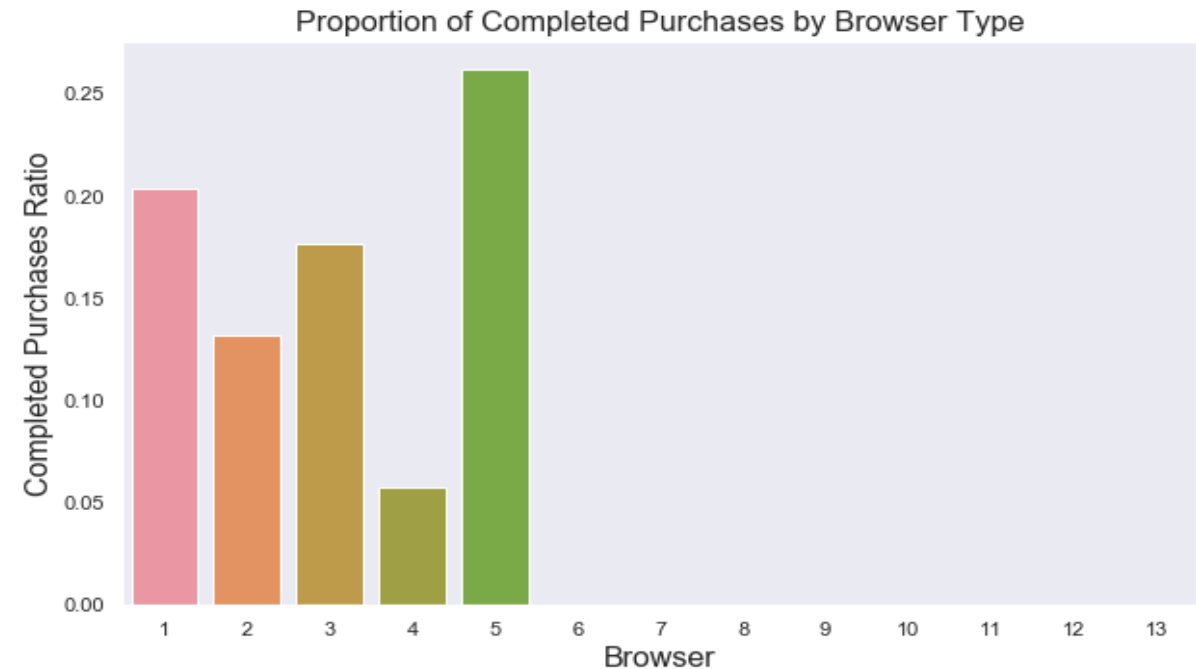
DATASET!

- From Kaggle: [Online Shopper's Intention](#)
 - Google Analytics data of various Turkish Websites

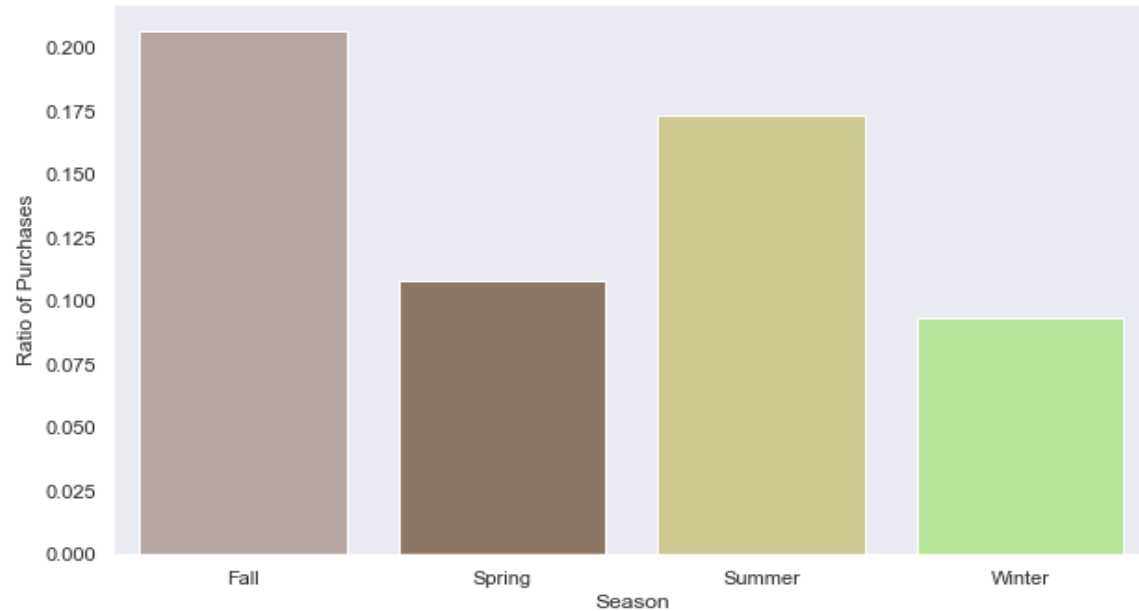


HOW ARE BROWSERS AFFECTING SALES?

- Browser 5 had the highest Purchase Ratio
- Browser 3 and 12 have high bounce rates:
 - Meaning the visitor didn't click on any other pages
 - UX/UI Designers should focus on compatibility of browser



Season vs Purchase Ratio



$H_0: \text{Purchase}_{\text{Fall}} = \text{Purchase}_{\text{Spring}} = \text{Purchase}_{\text{Summer}} = \text{Purchase}_{\text{Winter}}$

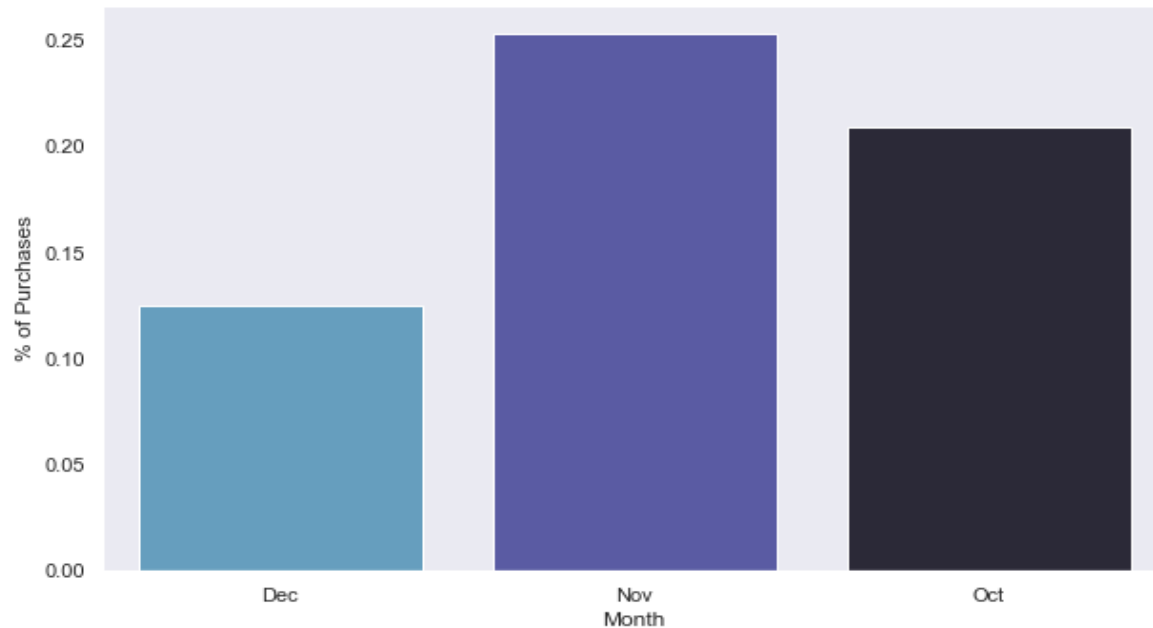
$H_a: \text{Purchase}_{\text{Fall}} \neq \text{Purchase}_{\text{Spring}} \neq \text{Purchase}_{\text{Summer}} \neq \text{Purchase}_{\text{Winter}}$

Chi-2 Test:
 Crit Value= 7.815
 Statistic = 816
 P-Value = 0.000
 Reject Null Hypothesis

SEASONALITY OF PURCHASE

- Fall season had the highest proportion of purchases
- In the fall, November had the highest proportion of purchases
- Turkish holidays in November: Republic Day, Religious holidays, Death Anniversary of the founder of Turkey

Fall Months Purchases



$H_0: \text{Purchase}_{\text{October}} = \text{Purchase}_{\text{November}} = \text{Purchase}_{\text{December}}$

$H_a: \text{Purchase}_{\text{October}} \neq \text{Purchase}_{\text{November}} \neq \text{Purchase}_{\text{December}}$

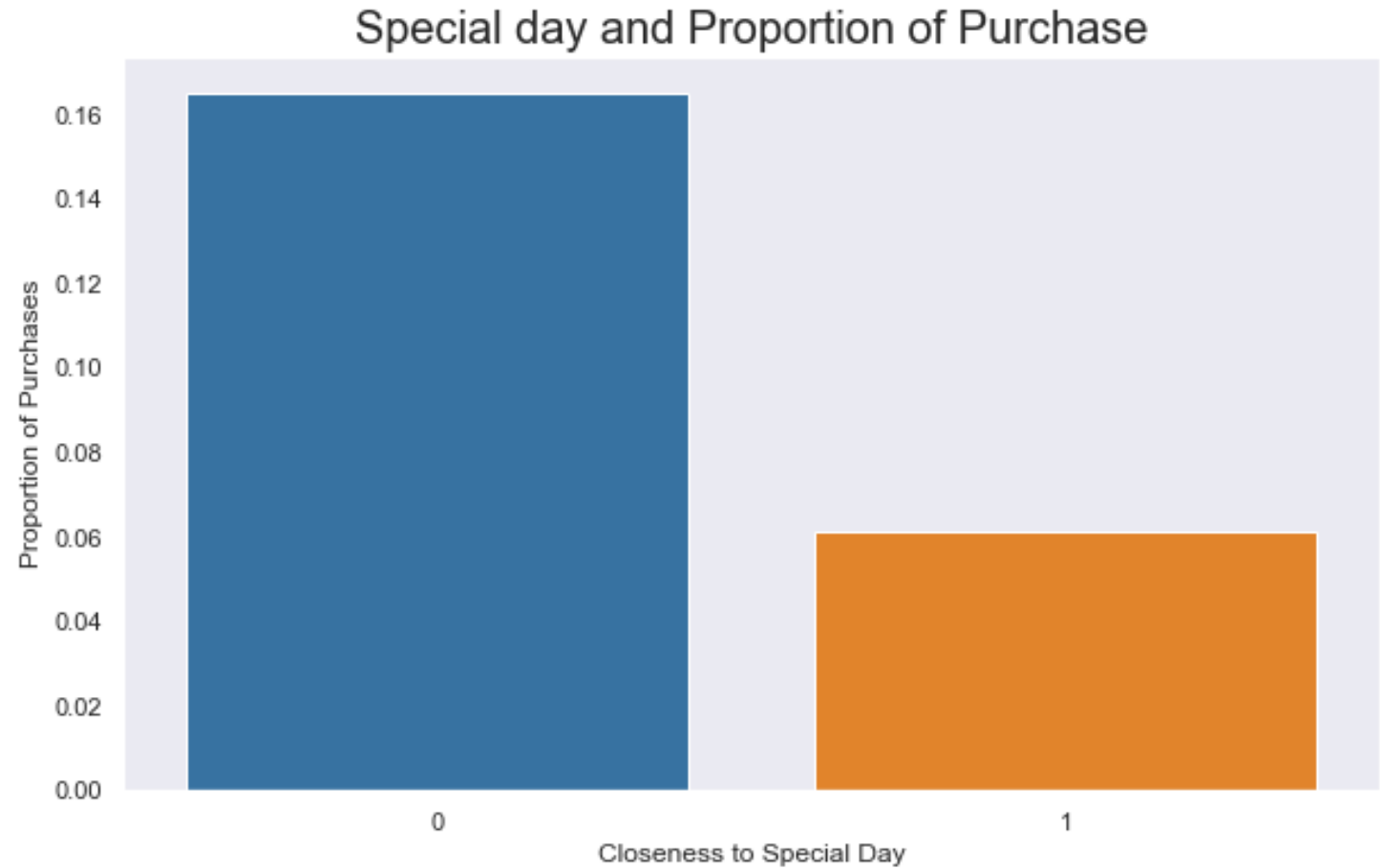
Chi-2 Test:
 Crit Value= 5.99
 Statistic = 110
 P-Value = 0.000
 Reject Null Hypothesis

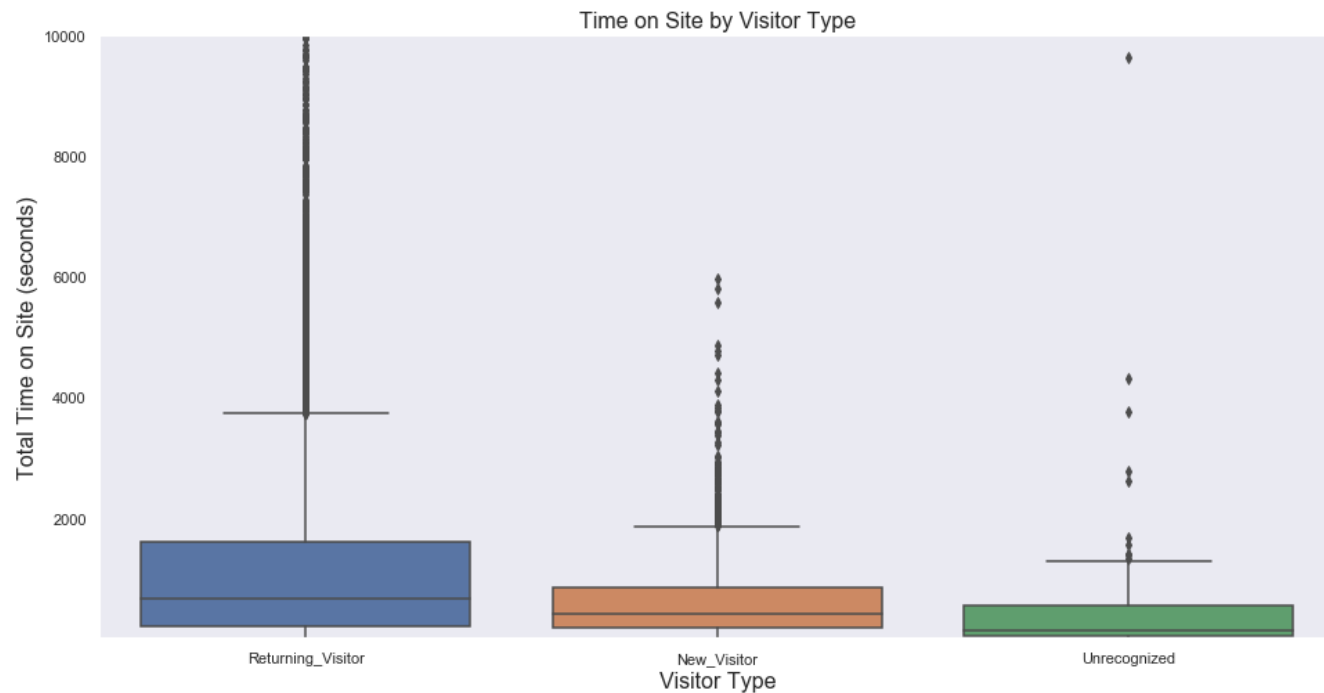
SPECIAL HOLIDAYS AND SALES!

- Significant difference between ratio of purchasing near
- Offer discounts near holidays

$$H_0: \text{Purchase}_{\text{Fall}} = \text{Purchase}_{\text{Spring}}$$

$$H_a: \text{Purchase}_{\text{Fall}} \neq \text{Purchase}_{\text{Spring}}$$

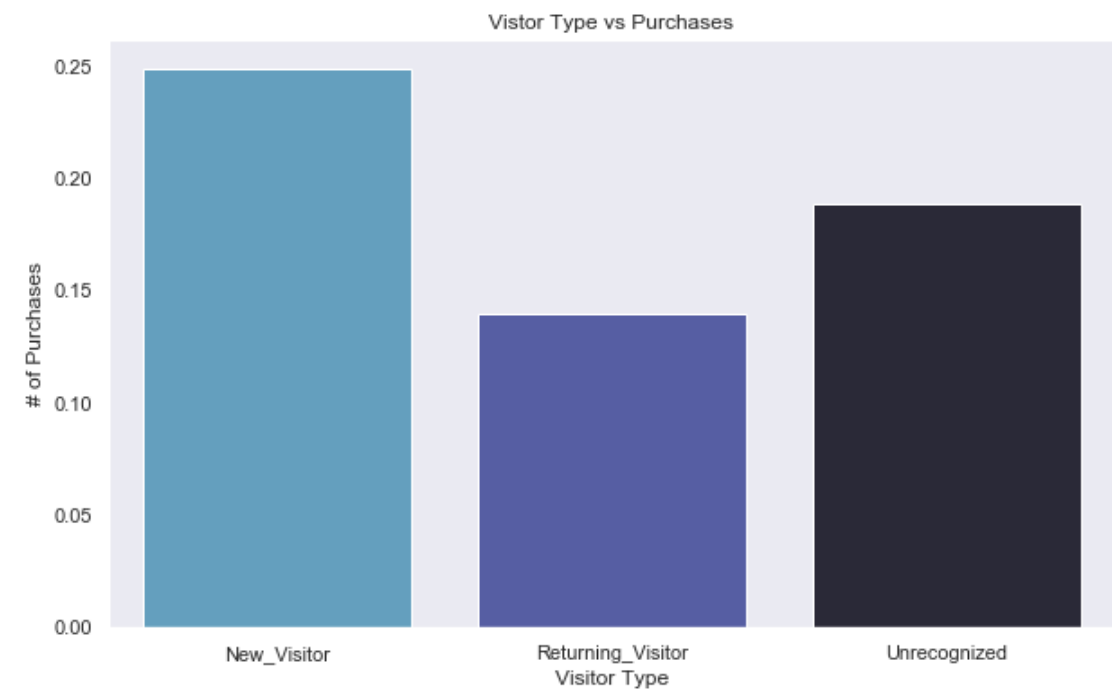




$H_0: Duration_{Returning} = Duration_{New} = Duration_{Unrecognized}$

$H_a: Duration_{Returning} \neq Duration_{New} \neq Duration_{Unrecognized}$

Anova Test:
 F-Stat = 81.7
 P-Value = 5.5×10^{-36}
 Reject Null Hypothesis



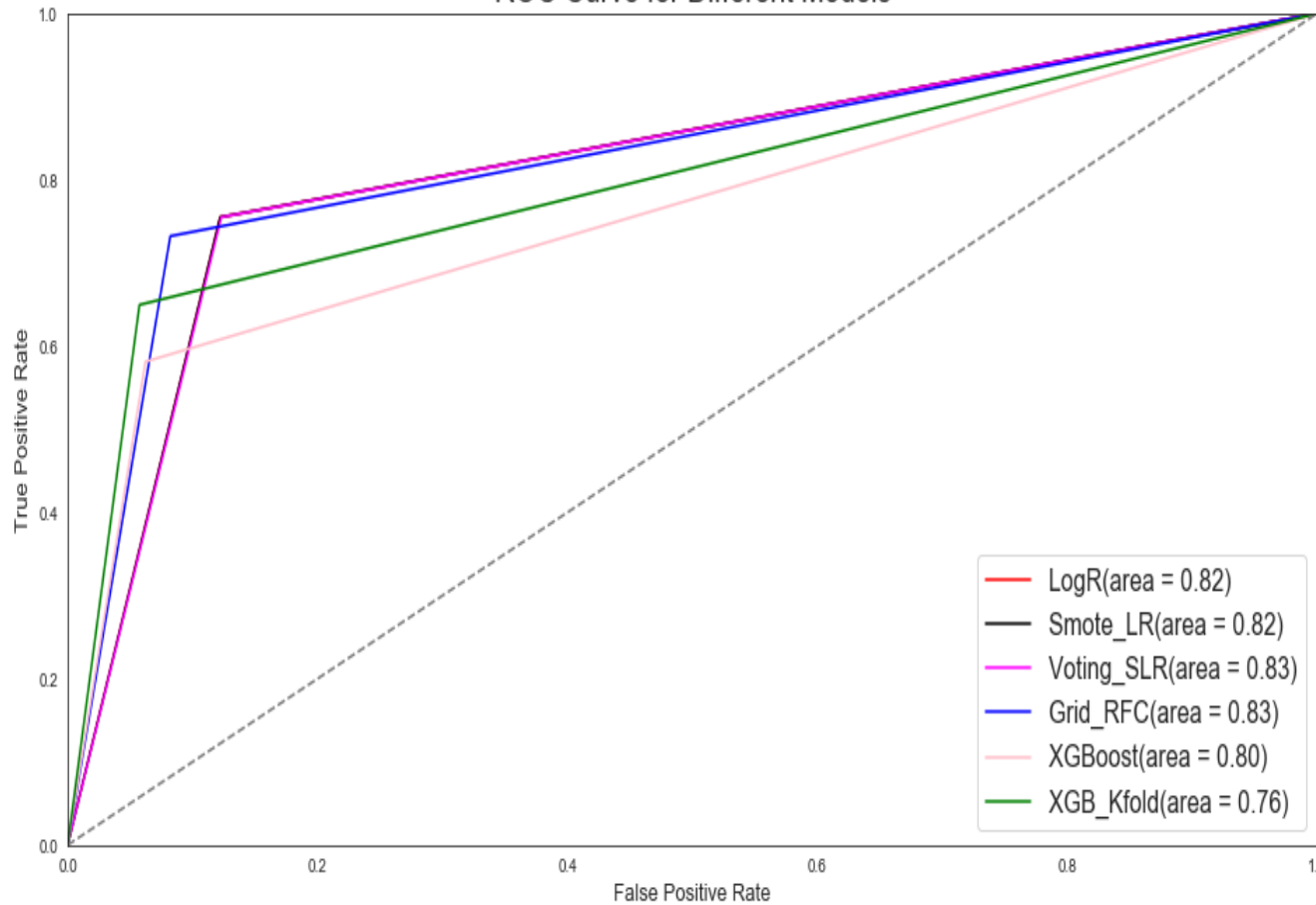
$H_0: Purchase_{Returning} = Purchase_{New} = Purchase_{Unrecognized}$

$H_a: Purchase_{Returning} \neq Purchase_{New} \neq Purchase_{Unrecognized}$

Chi-2 Test:
 Crit Value= 5.99
 Statistic = 135.3
 P-Value = 0.000
 Reject Null Hypothesis

- Returning Visitors spend more time on the website
- But they buy less product
- Can work on a reward system

ROC Curve for Different Models



Modeling Technique	Accuracy Score	F1 Score
Class Imbalance, Dominant Class	0.85	
Logistic Regression w/o Balancing	0.88	0.478
SMOTE Log Regression	0.86	0.607
Voting Classifier For SMOTE Log	0.86	0.606
XG Boost with K Fold	0.90	0.651
Random Forest w/ Grid Search	0.89	0.658

Hyper-Parameters:

Criterion: Gini
 Max_depth: 19
 Max_features: 8
 N_estimators: 800

MODELING

CONCLUSIONS AND RECOMMENDATIONS

Fix UI/UX design for compatibility of all browsers

Offer more seasonal products to increase revenue throughout the year

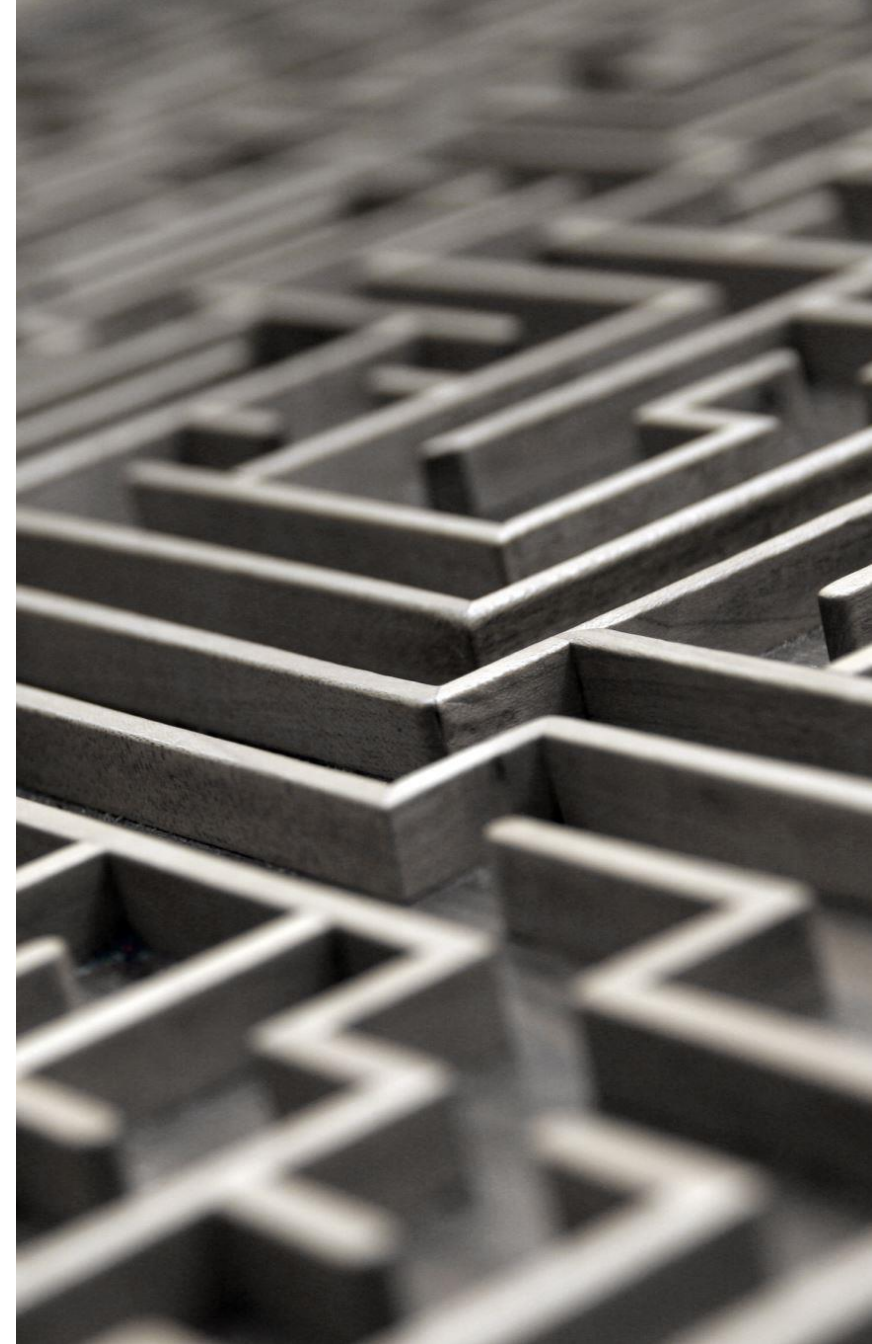
Offer discounts near special holidays, possibility of losing clients to other businesses

Offer loyalty programs to increase retention rates

Random Forest on a SMOTE sample gave us our best predictive model

LIMITATIONS

- Browser names not known
- GridSearchCV of our Random Forest Classifier gave us a max_depth of 19, which can lead to more overfitting
- Not knowing what specific websites the source of the data is coming from
- Google Analytics doesn't track New/Returning visitors anymore



**THANK
YOU!**

