

A grayscale photograph of a person from the chest up. They are wearing a light-colored, vertically striped shirt and a dark baseball cap. Their hands are visible; one hand holds a white smartphone horizontally, and the other hand holds a yellow pencil, pointing it towards the phone's screen. The person appears to be looking down at the device.

Predicting Time of Arrival for Food Delivery Service

Daniel Carpenter,
Sonaxy Mohanty,
Zachary Knepp

December 2022

Overview



Problem

Understand problem background

Data

Data preparation and cleansing
to perform modeling

Modeling

Overview and results of predictive
models

The Problem

MOTIVATION FOR CREATING PREDICTIVE MODEL



- 01.** This analysis aims to predict estimated delivery times for a food delivery service.
- 02.** With the increasing use of delivery systems, companies like Amazon, would want to know obstacles that may affect estimated delivery time.
- 03.** Providing accurate estimates to the customer will help manage expectations, which may lead to retained customers.

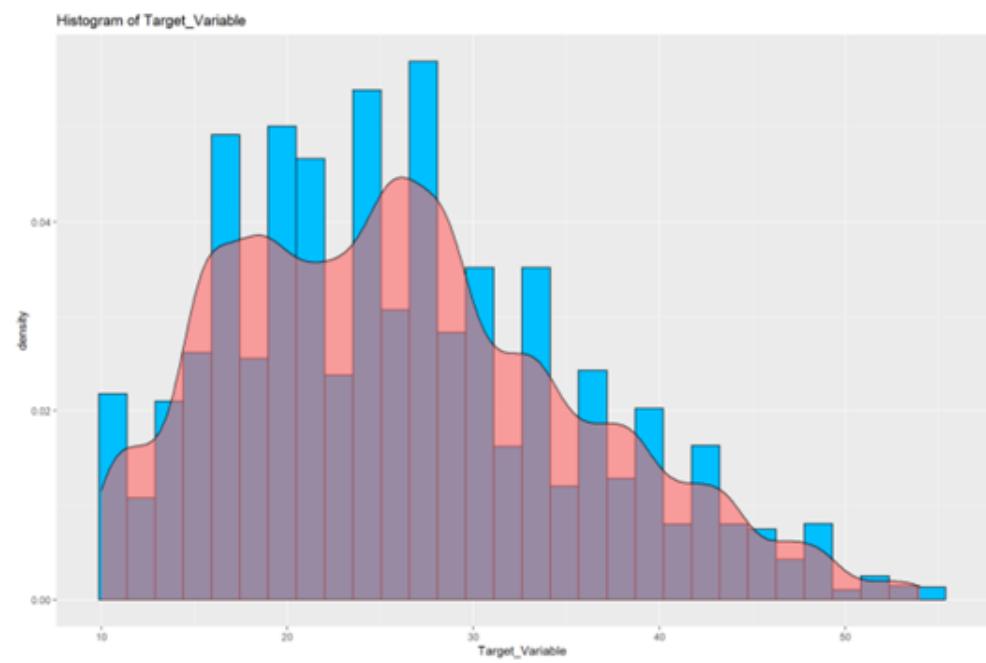
Data Prep.

Data Cleansing required for modeling

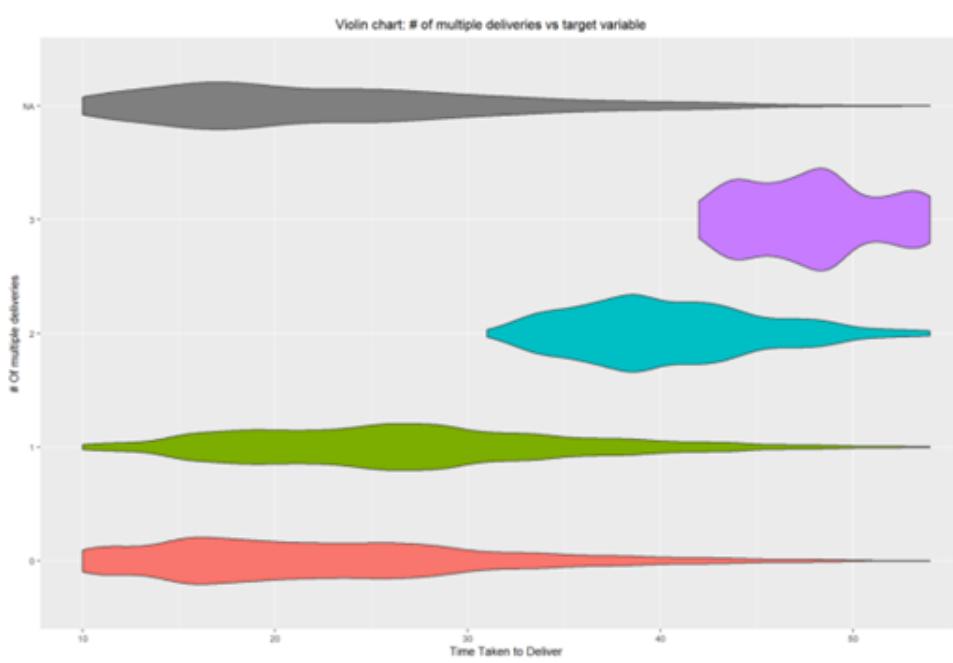
```
15  background: transparent;
16  box-sizing: border-box;
17  box-shadow: 0 15px 25px transparent;
18  border-radius: 10px;
19 }
20 .box h2{
21  margin: 0 0 30px;
22  padding: 0;
23  color: #fff;
24  text-align: center;
25 }
26 .box h3{
27  margin: 0 0 10px;
28  padding: 0;
29  color: #fff;
30  text-align: center;
31 }
32 .box .inputBox{
33 }
34 } position: relative;
35 .box .
```

- 01.** Combines and cleans 65,000 text files resembling each row of the data.
Includes 19 total features.
- 02.** Uses K-Nearest Neighbors (kNN) to impute missing factor data.
- 03.** Uses predictive-mean matching (PMM) to impute missing numeric data.
- 04.** Normalizes highly skewed distributions, including the target variable.
- 05.** No factor-lumping is required due to limited unique values.

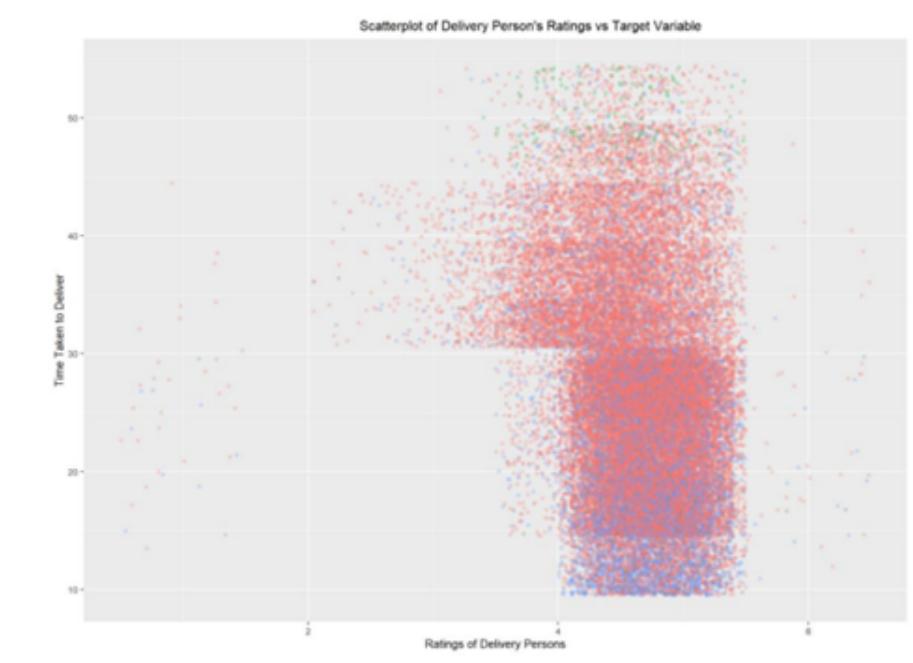
Data Exploration



**Target Variable:
Delivery Time**



**Multiple Deliveries
vs. Delivery Time**



Rating vs. Delivery Time

Modeling Overview

- 01.** Used various machine learning algorithms to predict the time to deliver orders.
- 02.** Goal: reduce the root-mean-squared error (RMSE) by comparing modeling performance.



OLS Results

A stepwise variable selection process is carried out to simplify the model without impacting much of the performance.

RMSE 01 0.786

Adj. R-Squared 02 0.551

MARS Results

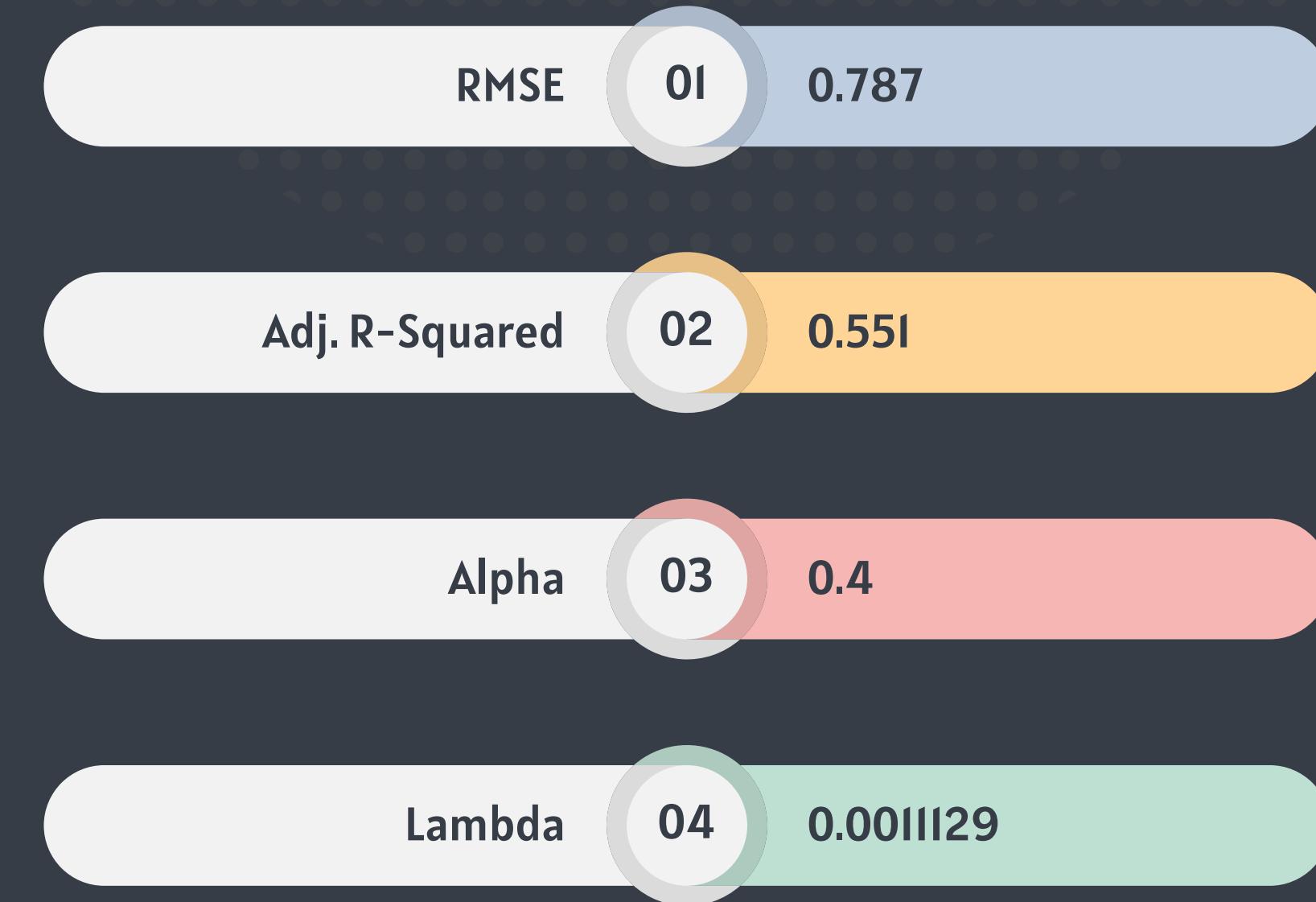
RMSE 01 0.761

Adj. R-Squared 02 0.580

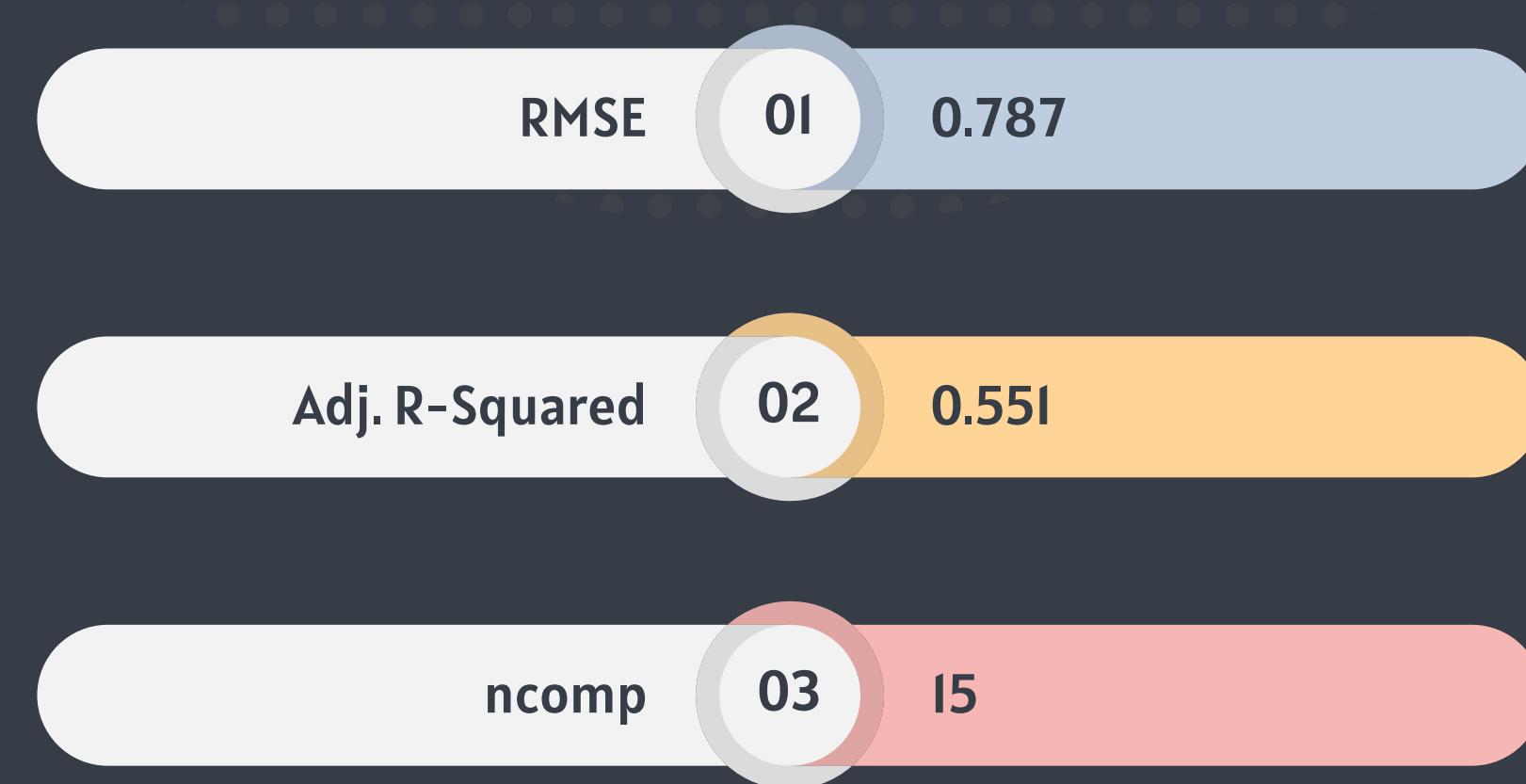
Degree 03 1

Prune 04 23

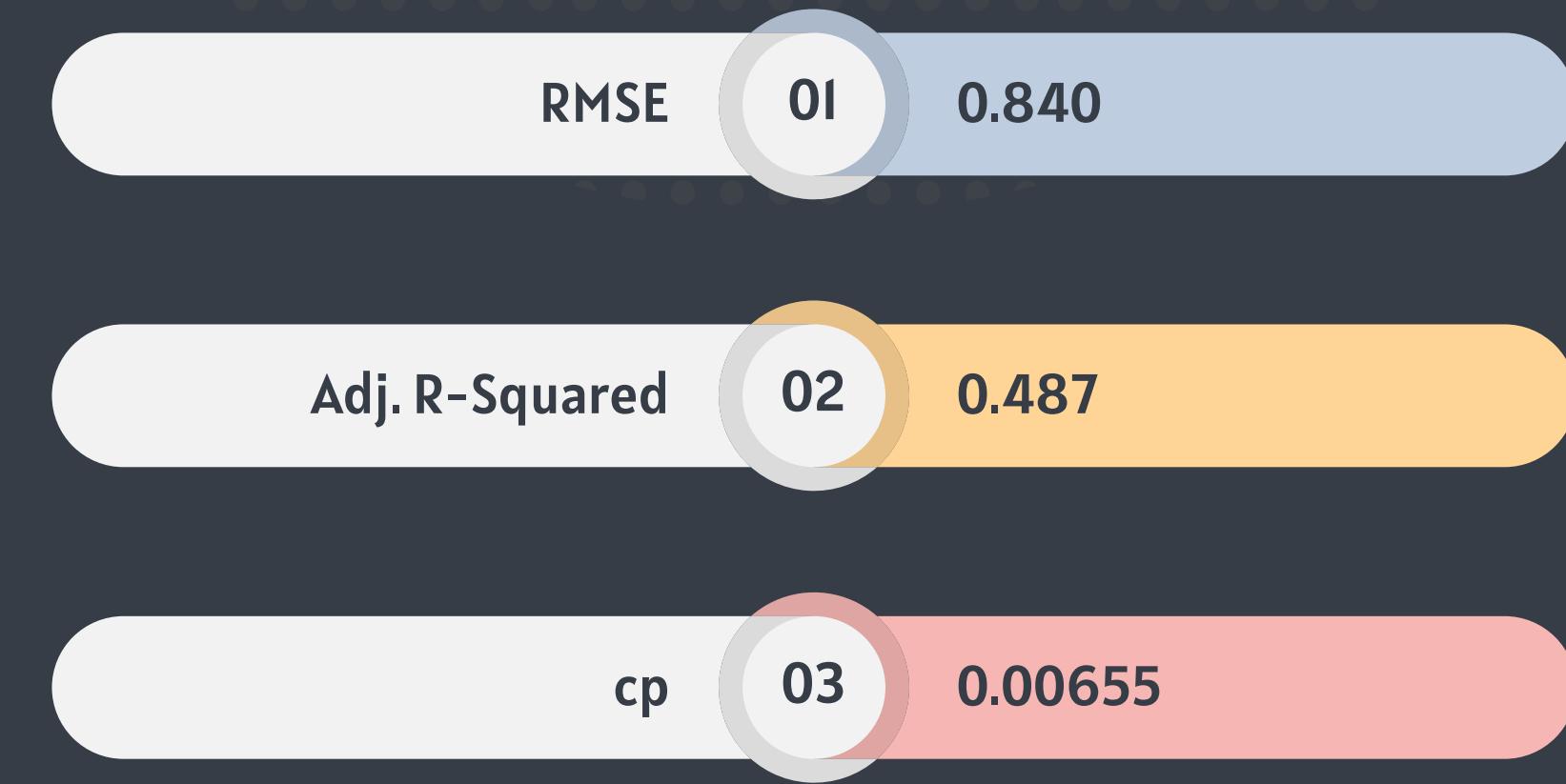
Elastic Net Results



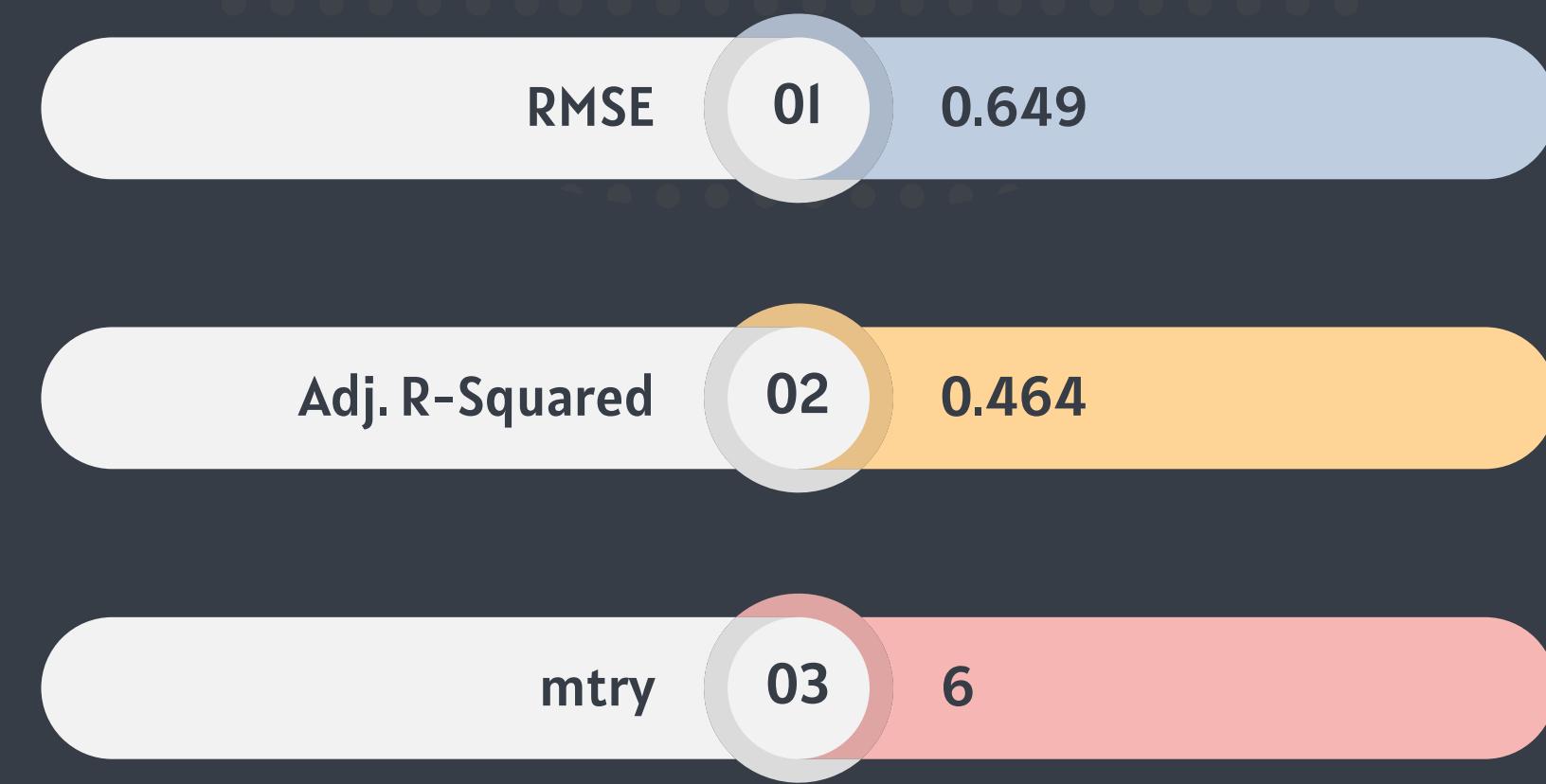
PCR Results



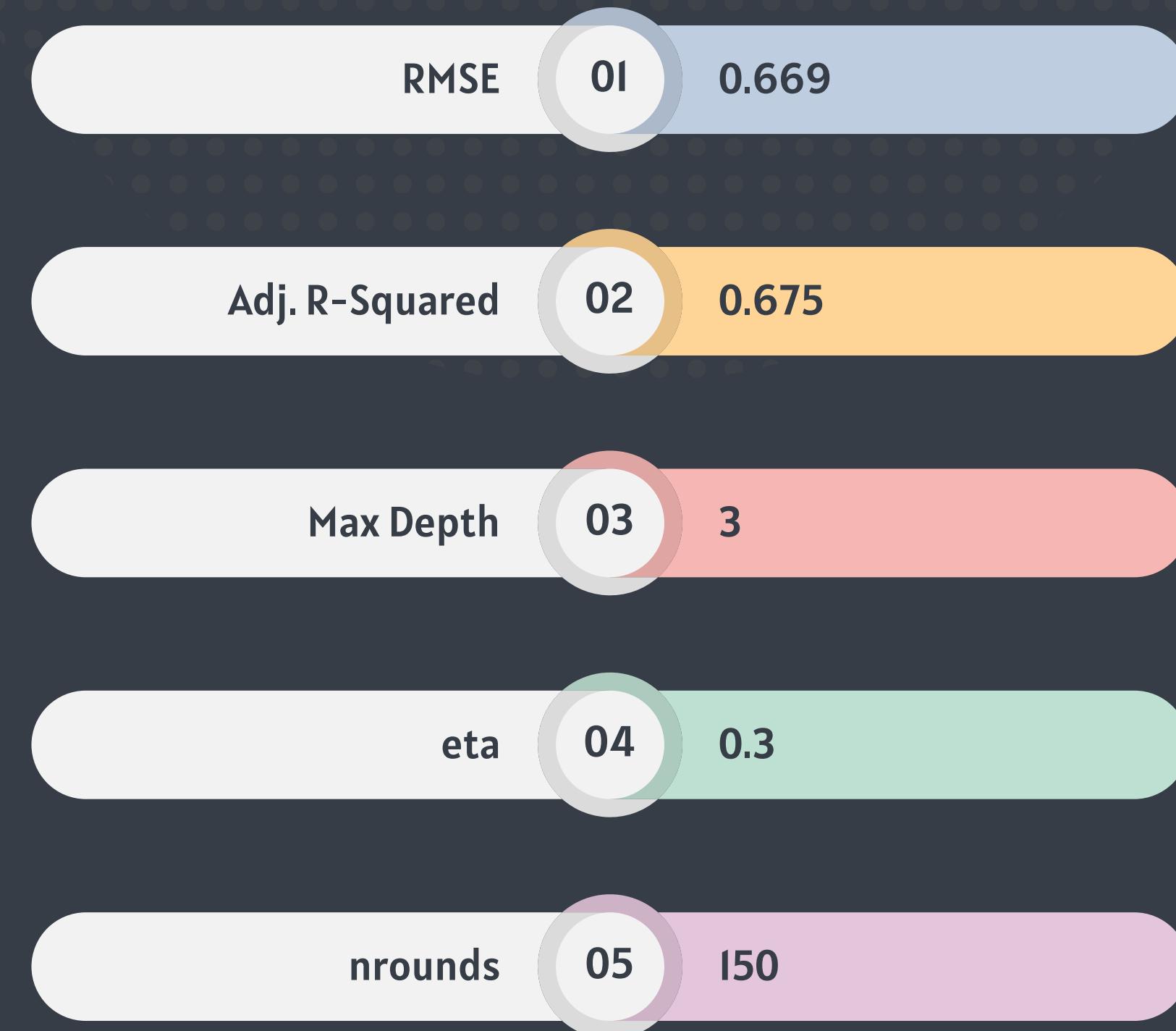
Decision Tree Results



Random Forest Results



Gradient Boosted Results



Model Conclusion

Model	Function or Package	Hyperparameters	RMSE	Adj. R-Squared
Random Forest	rf	mtry = 6	0.649	0.464
Gradient Boost	xgbTree	max_depth = 3 eta = 0.3 nrounds = 150	0.669	0.6751
MARS	caret and earth	Degree = 1 nprune = 23	0.761	0.580
OLS	lm	N/A	0.786	0.551
Elastic Net	caret and elasticnet	Alpha = 0.4 Lambda = 0.0011129	0.787	0.551
PCR	pca	ncomp = 15	0.787	0.551
Decision Tree	rpart	cp = 0.0065568	0.840	0.487

Summary



Delivery Problem



Data Cleansing



Predictive Models

