

Project 2:

한정판 신발 리셀가 예측

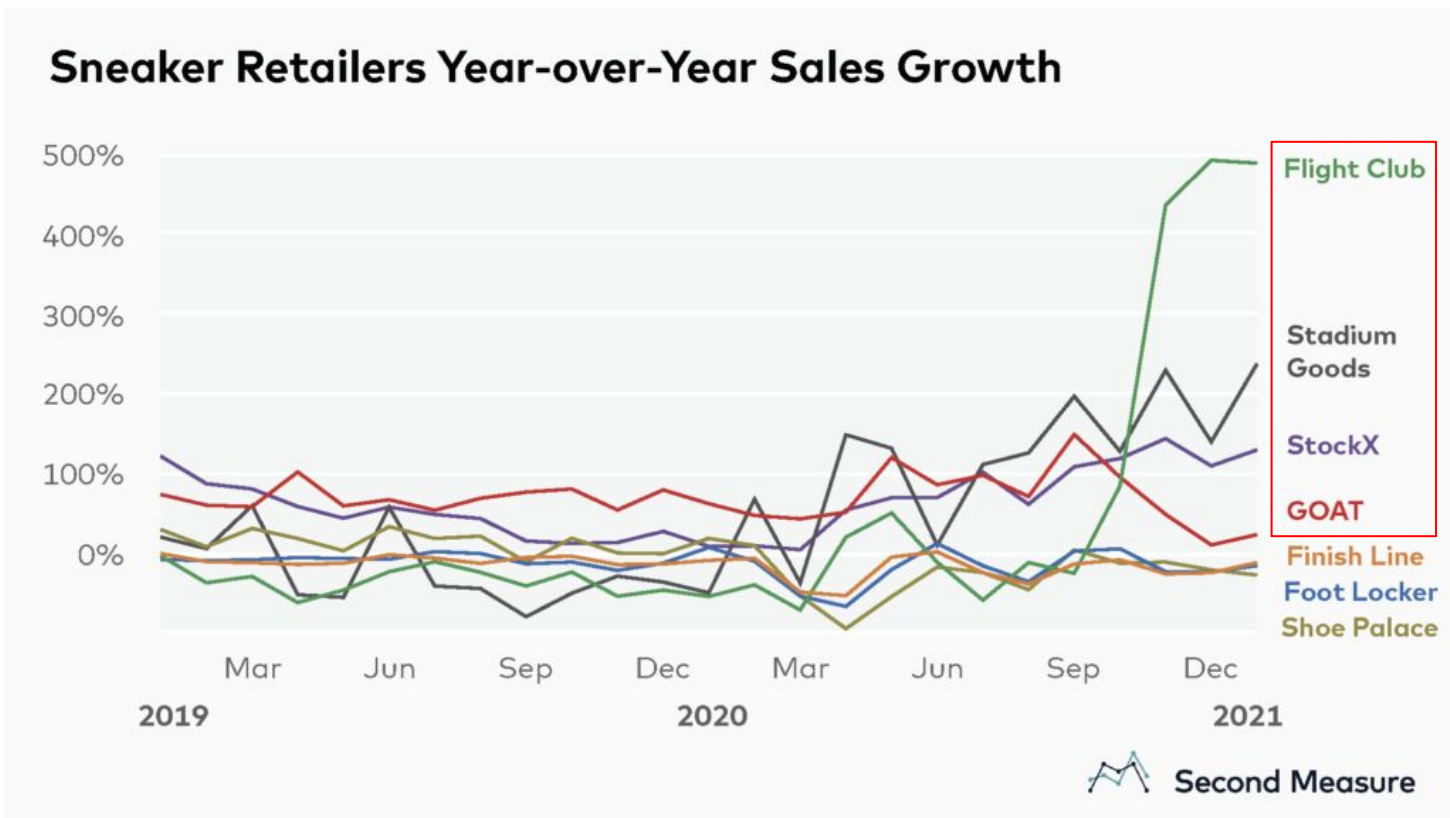
AI_10_임규현

Introduction

- 커져가고 있는 신발 리셀 마켓



Introduction



Introduction

- 어떤 요소가 리셀가를 예측하는데 도움이 되는가?
- 시간이 지날 수록 리셀가는 올라가는가?
- 매수/매도 하기 좋은 시기가 존재하는가?
- 사이즈에 따라 리셀가가 변하는가?

About Data

- 10 만개의 StockX 의 신발 거래 데이터 (2017년 9월~ 2019년 4월)
- 총 53가지의 신발 (아디다스 이지, 나이키 오프화이트)
- Order Date, Release Date, Sneaker Name, Sale Price, Retail Price, Buyer Region, Shoe Size, Brand
 - → Days_Since_Release, Sneaker Name, Shoe Size, Buyer Region (Top 5), Color



stockX



View All Sales

SIZE	SALE PRICE	DATE	TIME
10	\$890	Wednesday, October 16, 2019	5:49 PM EST
8.5	\$1,174	Wednesday, October 16, 2019	5:47 PM EST
11	\$729	Wednesday, October 16, 2019	5:47 PM EST
10	\$870	Wednesday, October 16, 2019	5:46 PM EST
8	\$1,183	Wednesday, October 16, 2019	5:44 PM EST

12 MONTH HISTORICAL

OF SALES

3785

PRICE PREMIUM

(OVER ORIGINAL RETAIL PRICE)

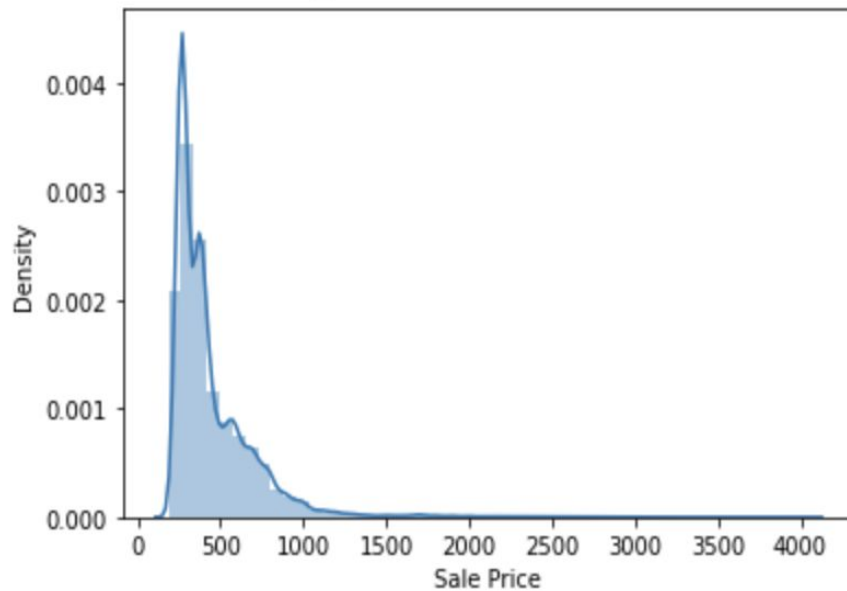
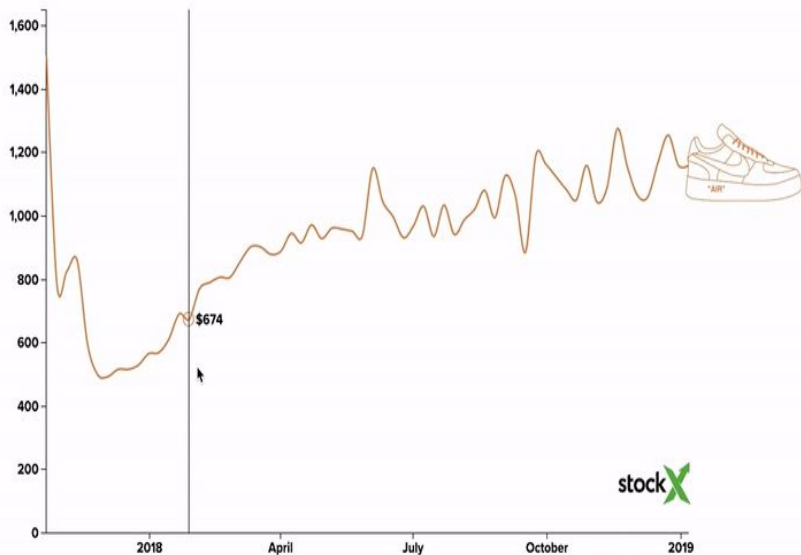
312.0%

AVERAGE SALE PRICE

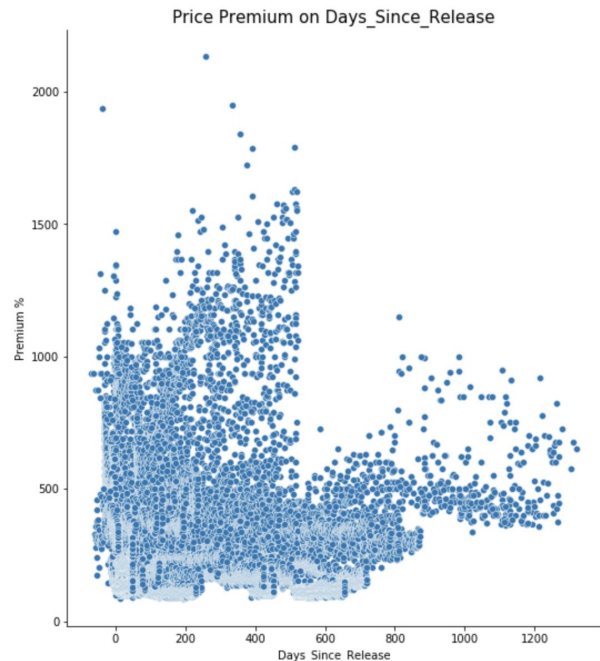
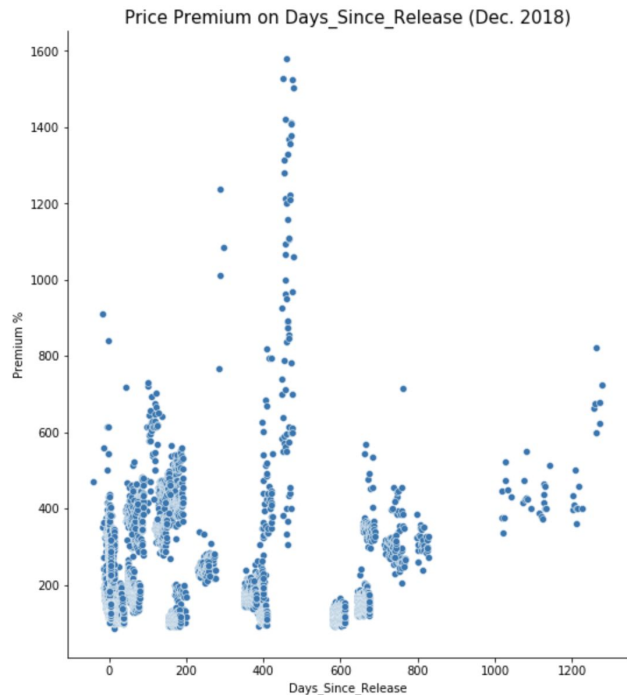
\$996

Problem Statement

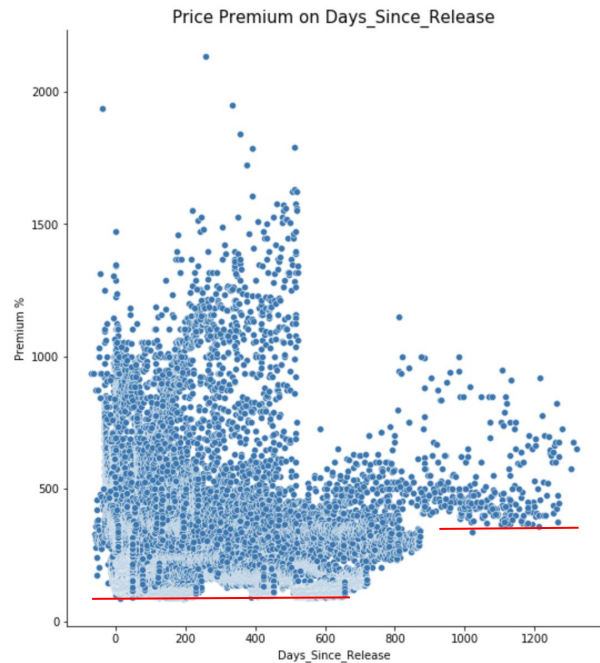
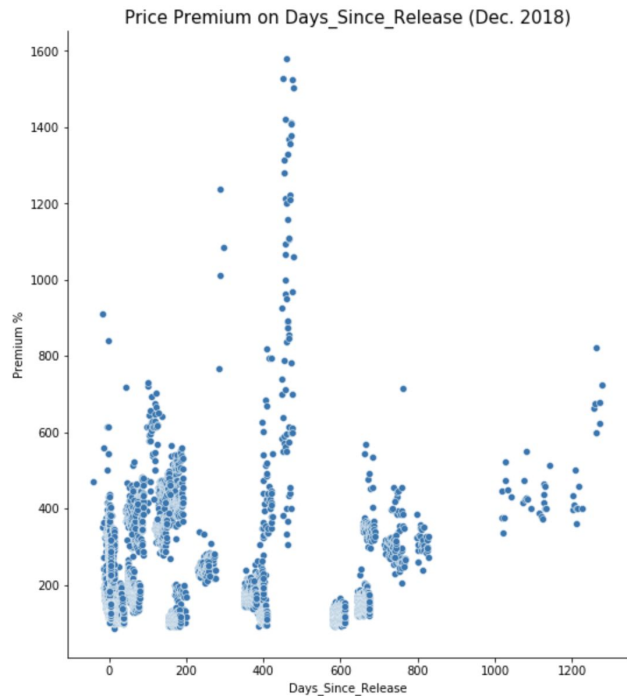
- 신발의 리셀가 (Sale Price 예측)
- 회귀 모델로 접근



Premium vs. Time

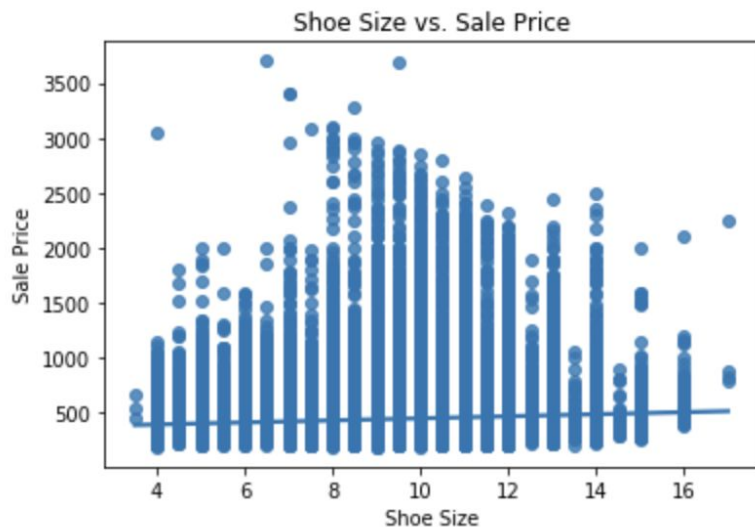


Premium vs. Time



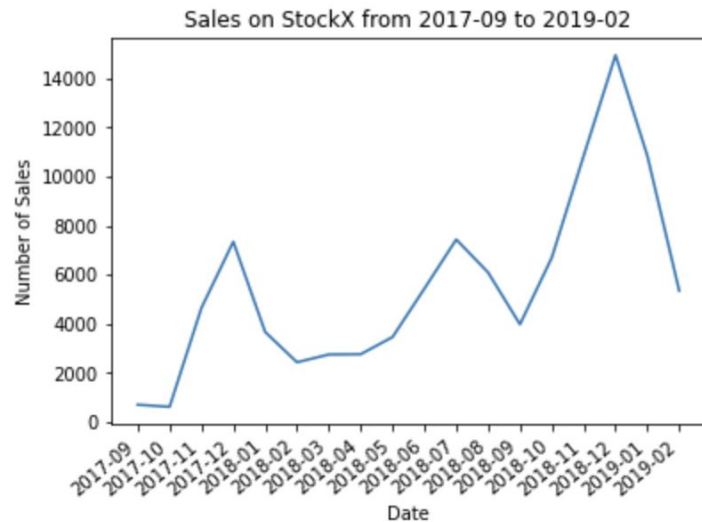
Size vs. Sale Price

- 신발 사이즈가 리셀가에 끼치는 영향이 있는가?




EDA

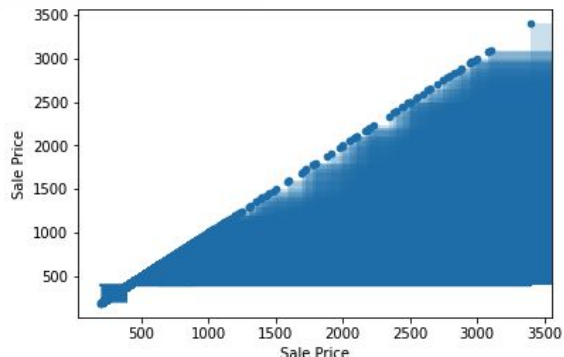
- 결측치 점검 / 제거
- 데이터 Feature 변형 및 추가
 - Sneaker Name
 - 53 -> 10
 - Color
 - Days_Since_Release
 - Buyer Region
 - 50 -> 6
 - Number of Sales
- 학습시킬 데이터 선정
 - 가장 거래량이 많은 달을 기준 (**2018년 12월**)
 - 약 12,000 개의 거래 데이터로 훈련



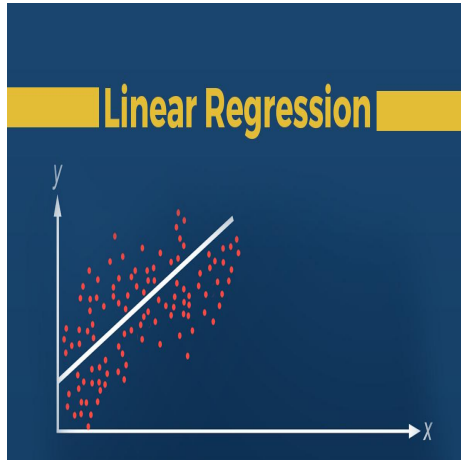
Model Selection

- Baseline Model (Sale Price)
 - 목적 : 최소한의 성능을 나타내는 것
 - 리셀가의 **평균**이 기준

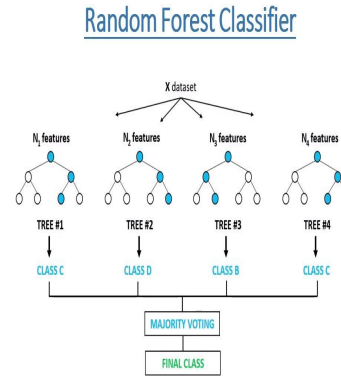
	Metric	Score	
0	MSE	41742.398169	
1	MAE	132.516611	
2	RMSE	204.309565	
3	R2	0.000000	



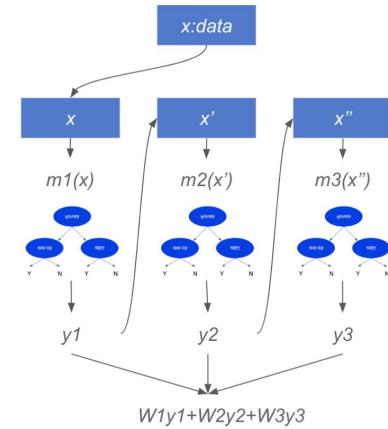
Model Selection



- R- Squared: 0.76

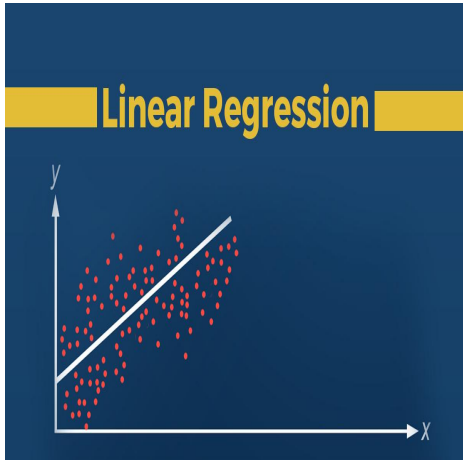


- R- Squared: 0.08



- R- Squared: 0.10

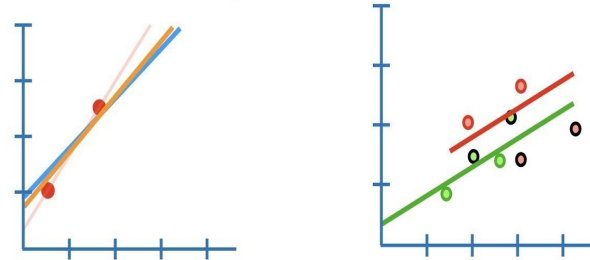
Model Selection



- R-Squared: 0.76
- MAE: 39.30
- MSE: 4695.33



Lasso Regression....



...Clearly Explained!!!

- R-Squared: 0.77
- MAE: 39.30
- MSE: 4694.69
- CV-Score: 0.82

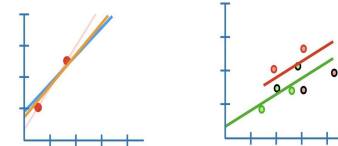
Model Selection

Baseline Model

	Metric	Score
0	MSE	41742.398169
1	MAE	132.516611
2	RMSE	204.309565
3	R2	0.000000

VS

Lasso Regression....



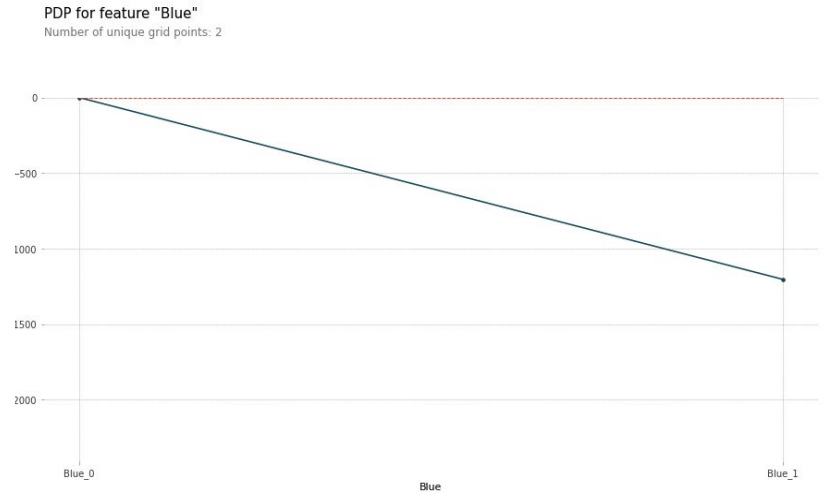
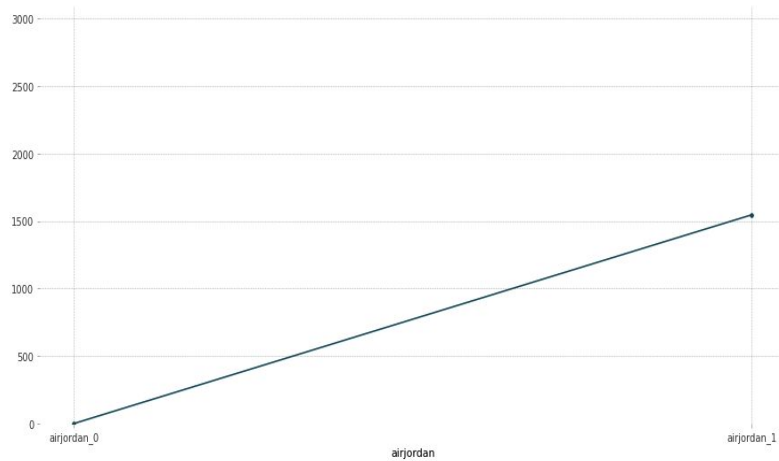
...Clearly Explained!!!

- MSE: 4694.69
- MAE: 39.30
- CV-Score: 0.82
- R-Squared: 0.77

Feature Importance (Coefficients)



Interpreting Coefficients (PDP Plot)



Prediction



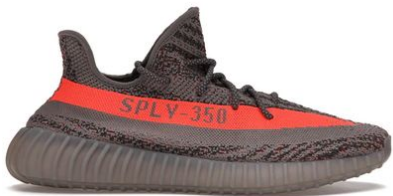
- Nike Air Jordan 1 High “Travis Scott”

- Retail : \$175, Color: Black/Brown, Days_Since_Released = 1936, Size = 11, State = New York, # Sold = 341

- Result: \$2,645 (Sale Price) vs. \$2,082 (actual)

- 실제 거래값과 27% 차이

- Adidas Yeezy Boost 350 V2 “Beluga Reflective”



- Retail: \$240, Color: Grey/Orange, Days_Since_Released = 25, Size = 10, State = California, # Sold = 1736

- Result: \$367 vs. \$328 (actual)

- 실제 거래값과 12% 정도 차이