

Section2 Project

AI_11_신성윤

Online Store Dataset(Olist)

shipping_limit_date	price	freight_value	payment_sequential	payment_type	payment_installments	...	product_name_lenght	product_description_lenght
2017-09-19 09:45:35	58.90	13.29	1.0	credit_card	2.0	...	58.0	598.0
2017-05-03 11:05:13	239.90	19.93	1.0	credit_card	3.0	...	56.0	239.0
2018-01-18 14:48:30	199.00	17.87	1.0	credit_card	5.0	...	59.0	695.0
2018-08-15 10:10:18	12.99	12.79	1.0	credit_card	2.0	...	42.0	480.0
2017-02-13 13:57:51	199.90	18.14	1.0	credit_card	3.0	...	59.0	409.0

- Olist의 브라질 전자상거래 공개 데이터셋
 - 약 10만건의 데이터셋

프로젝트 목표

평점에 영향을 주는 요소들을
파악해

평점이 낮은 판매자들이
평점 관리를 잘 할수 있도록
도와준다.

TABLE OF CONTENTS

01

데이터 선정이유
및
문제정의

02

모델작성전
평점관련 시각화

03

모델
학습/예측/평가

04

시각화 및 결론



01

데이터
선택 이유

데이터 선정 이유



데이터 선정이유

본래 유저 행동데이터를 얻으려
했으나 얻는데 제약이 있기
때문에
유사한 데이터를 선정



문제정의

배달의민족 관련기사에서
자영업자들이 평점에
무척 민감하다는글
확인

타겟 선정

타겟

“Review Score”



기준모델

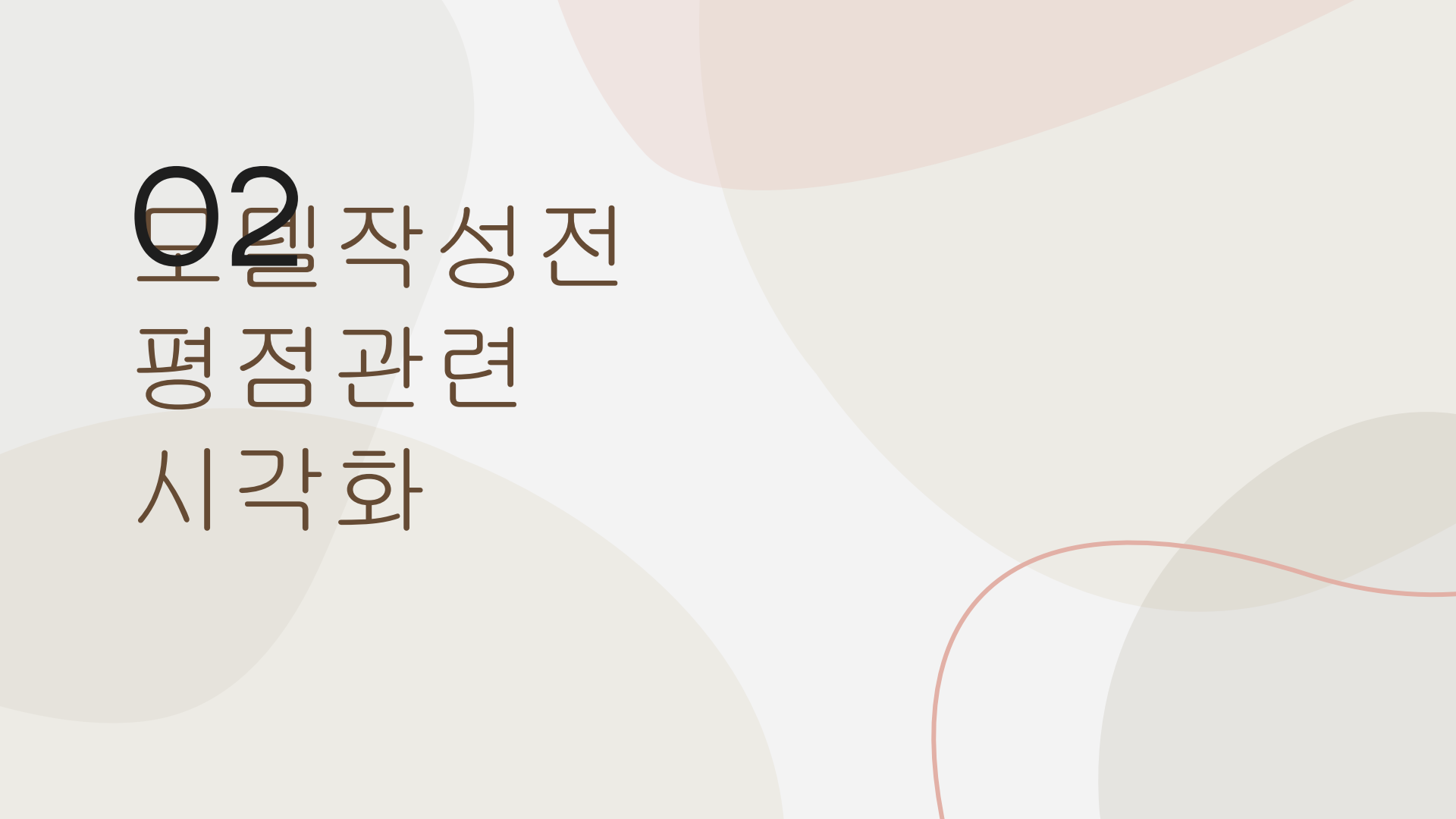
“Review Score”
최빈값



평가지표

“Roc-Auc”



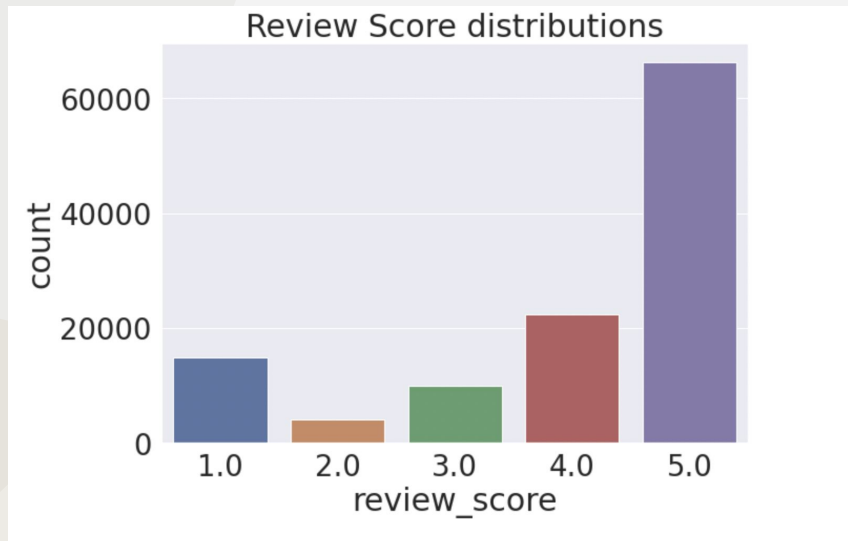


02 모델 작성 전

평점 관련

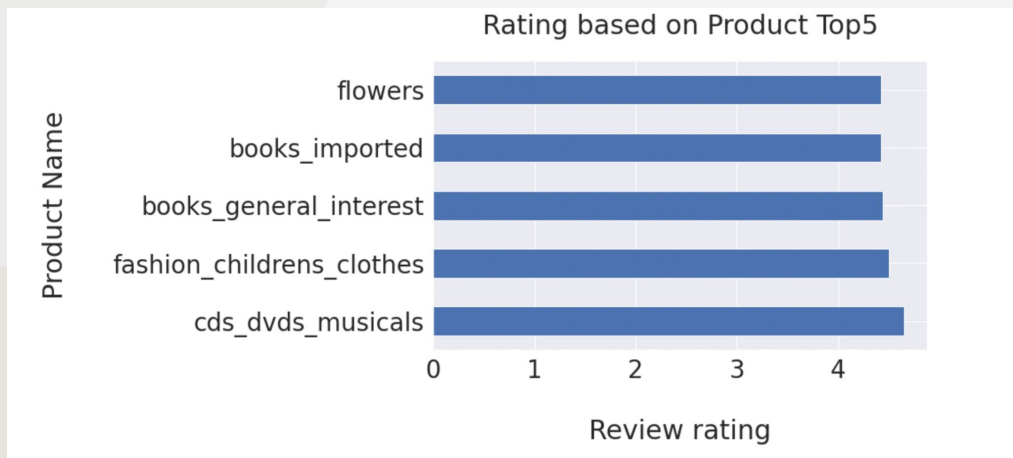
시각화

평점 분포



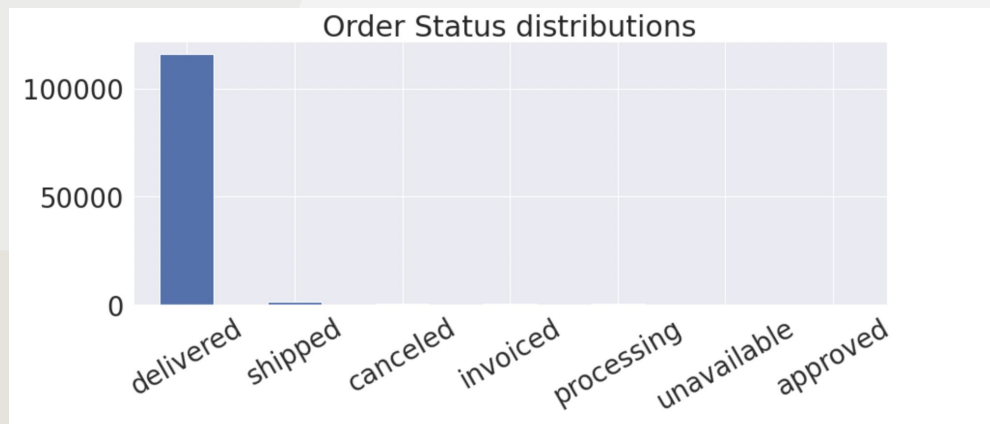
- 5점(최대) 과반수 이상
- 1~3점의 비율도 합치면
상당함

제품 평점 평균Top5



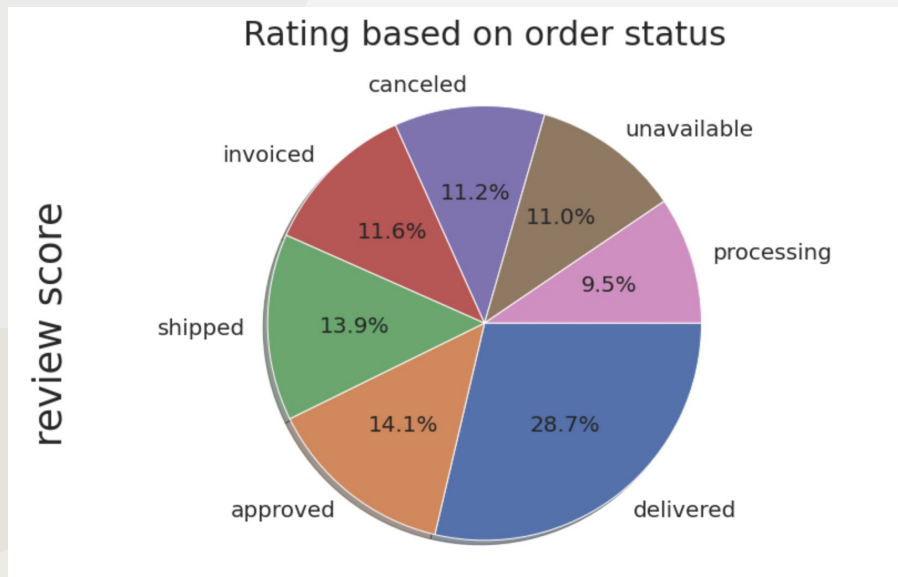
- 제품의 평점중 특별히
높고 낮은것이 있는지
확인

배송현황



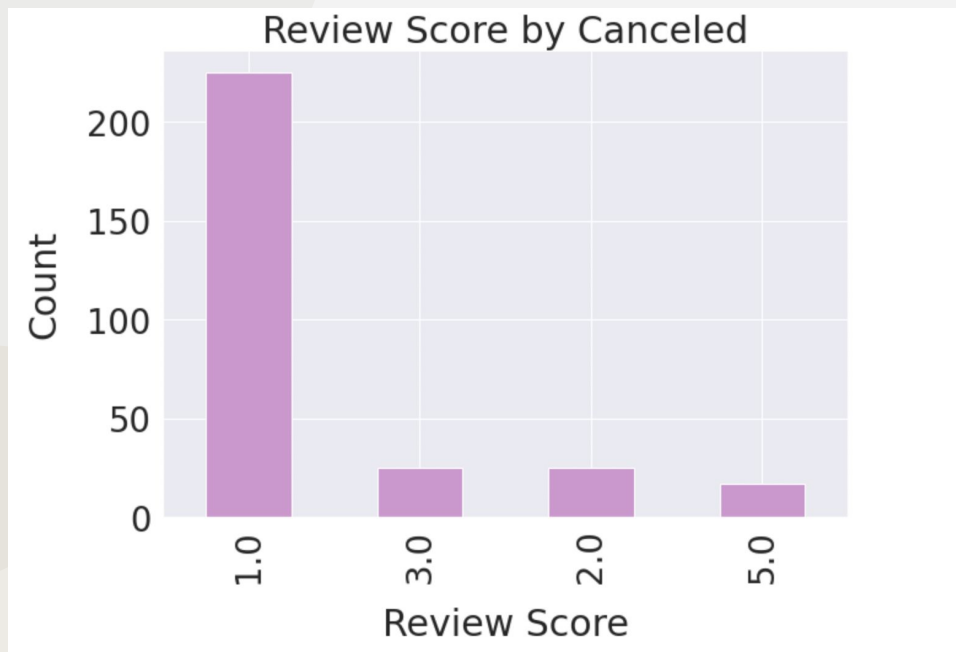
- 배송에 차질이 자주 발생하는지 확인
- 배송에 문제가 있을 경우 평점에 영향이 있을 가능성 높다.

배송현황에 따른 평점현황



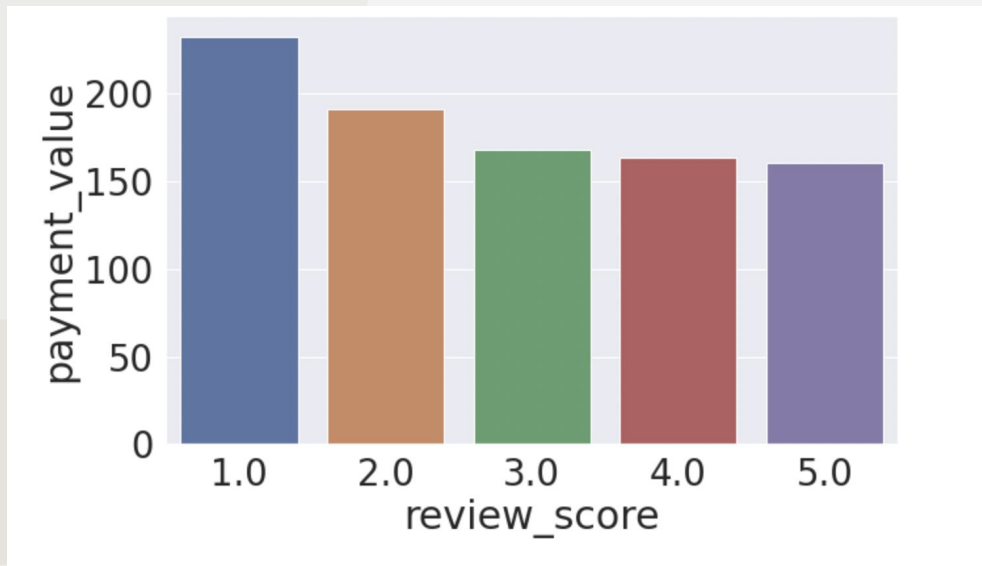
- 배송의 대부분은
차질없이 완료상태
- “주문취소,” “구매불가,”
“주문진행중”은 평점에
민감함

“주문취소”에 따른 평점



- 주문취소시 평점 1점
받을 가능성은 굉장히
높다.

지불비용에 따른 평점



- 지불비용이 높으면
평점이 낮다.
- 그러나 크게 차이나지
않는다.

질문



데이터 누수 여부

학습/검증/테스트 데이터를 잘 나눠주었으며 테스트 데이터에 스케일링 등 특별한 작업을 하지 않았다.



모델의 유용성 및 한계

과적합을 방지해주며 대용량 데이터처리에 좋지만 메모리 소모량이 크고 훈련 데이터를 추가해도 성능개선이 어렵다.

03
모델

학습 / 예측 / 평
가

모델 비교

Logistic Regression

AUC Score : 0.61
모델 정확도 : 0.59
기준모델 정확도 : 0.75



RandomForest Classifier

AUC Score : 0.81
모델 정확도 : 0.84
기준모델 정확도 : 0.75



XGBoost Classifier

AUC Score : 0.74
모델 정확도 : 0.74
기준모델 정확도 : 0.75



질문



베이스라인과 비교

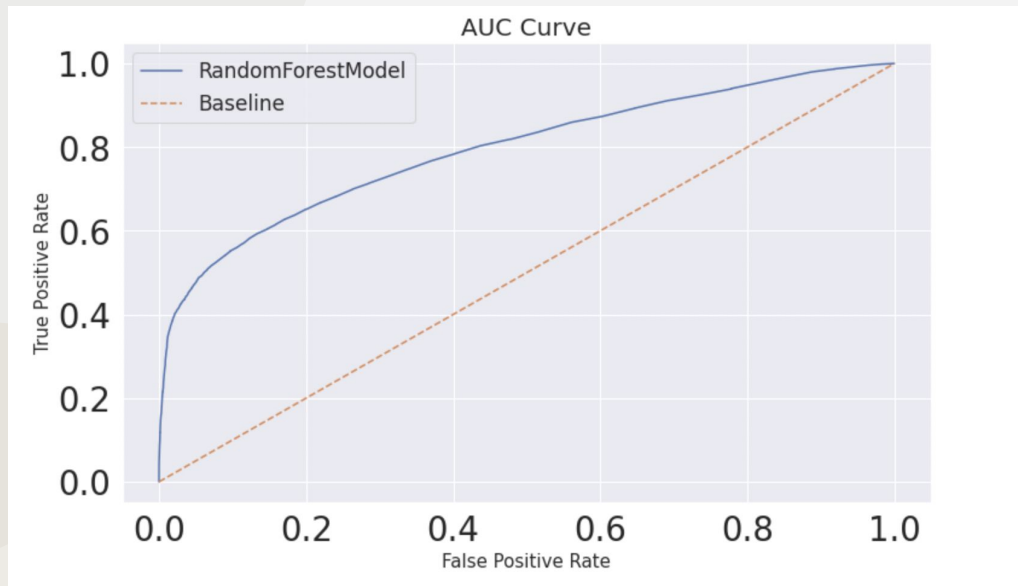
베이스라인은 정확도: 0.75
만든 모델은 정확도: 0.84



모델 성능 개선 방법

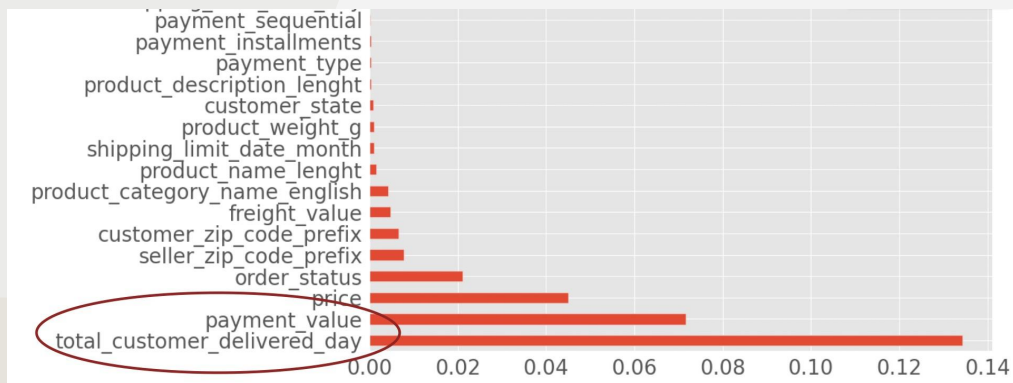
하이퍼파라미터 튜닝 (x)
결측치 제거
데이터 균형 작업

평가지표 : Roc-Auc



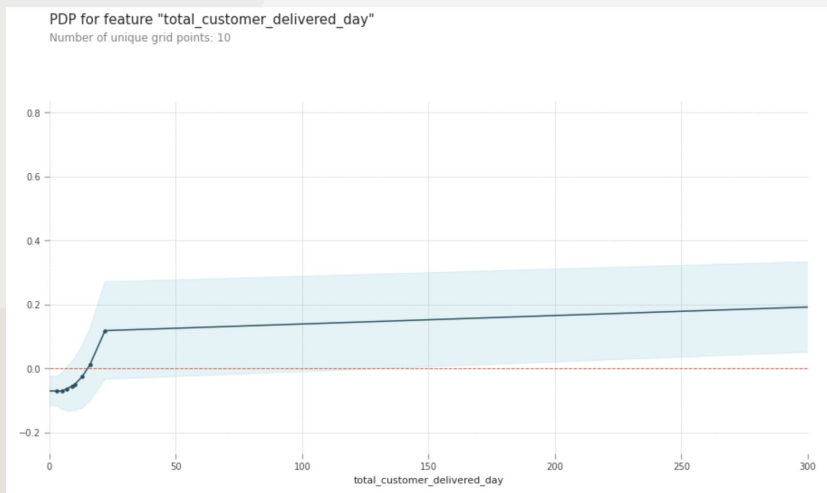
- 좌측 상단(1.0)으로 갈수록 성능이 좋음.
- TPR(예측값: 나쁨 / 실제값: 나쁨)
- FPR(예측값: 나쁨 / 실제값: 좋음)

평점에 영향을 주는 요소들



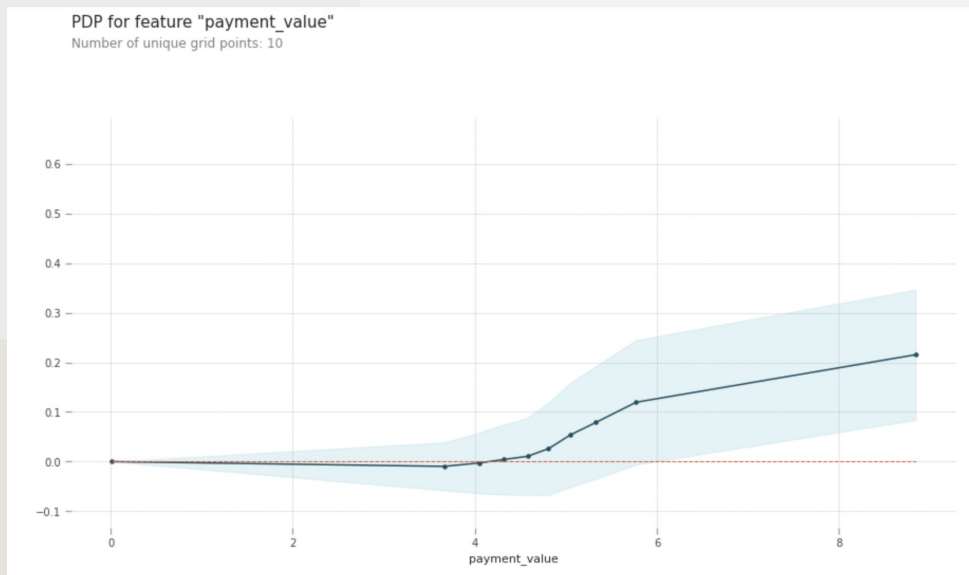
- 총 배송소요일
- 지불가격

1. 총 배송 소요일



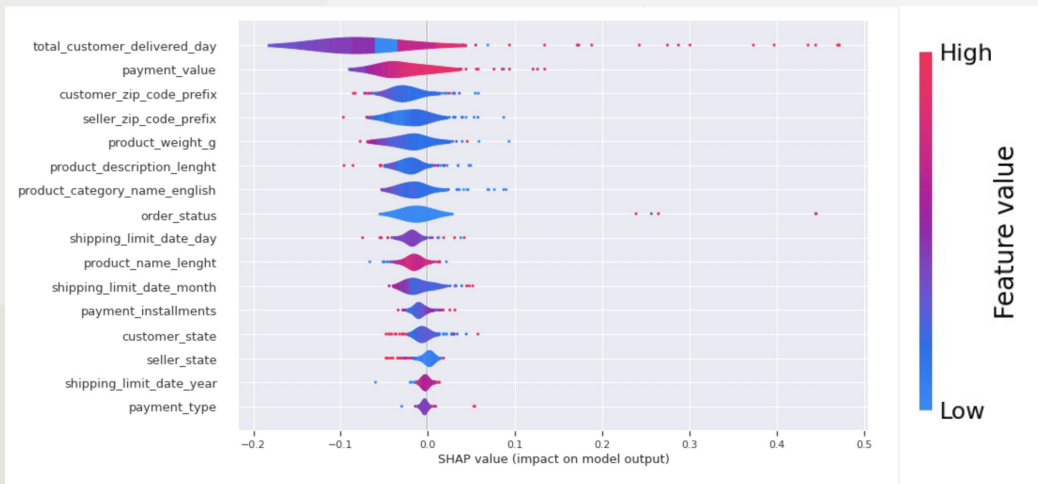
- 총 배송소요일이
높을수록 그래프가
상승한다.(평점이
안좋아진다.)

2. 지불 가격



- 지불가격이 높을수록
그래프가
상승한다.(평점이
안좋아진다.)

각 변수들 분포 확인



- 총 배송소요일 및
지불가격의 분포가 제일
넓음
- 값들이 작아질수록
(파란색) 평점이
좋아진다.



04

프로젝트

결론

평점이 낮은 사람들에게
“배송기간”, “지불가격” 에 대해
더 낮을수록 평점에 유리하다고
팁을 줄 수 있다.

6699

The background features several overlapping circles and a thin, wavy line in muted colors like light pink, beige, and grey. The text is centered in a clean, black, sans-serif font.

감사합니다.