



# 신용카드 승인 예측

MANGO MANGO

응웬탄텅  
이정호  
조진목



# 목차

**1. 배경**

**2. 데이터 수집**

**3. EDA**

**4. 데이터 전처리**

**5. 모델링 및 평가**

**6. 한계 및 결론**





# 데이터 수집 - Kaggle



```
Data columns (total 21 columns):
#      Column      Non-Null Count  Dtype
---  -
0      Applicant_ID  25128 non-null    int64
1      Applicant_Gender 25128 non-null    object
2      Owned_Car        25128 non-null    int64
3      Owned_Realty     25128 non-null    int64
4      Total_Children   25128 non-null    int64
5      Total_Income     25128 non-null    int64
6      Income_Type      25128 non-null    object
7      Education_Type   25128 non-null    object
8      Family_Status    25128 non-null    object
9      Housing_Type     25128 non-null    object
10     Owned_Mobile_Phone 25128 non-null    int64
11     Owned_Work_Phone  25128 non-null    int64
12     Owned_Phone       25128 non-null    int64
13     Owned_Email       25128 non-null    int64
14     Job_Title         25128 non-null    object
15     Total_Family_Members 25128 non-null    int64
16     Applicant_Age     25128 non-null    int64
17     Years_of_Working  25128 non-null    int64
18     Total_Bad_Debt    25128 non-null    int64
19     Total_Good_Debt   25128 non-null    int64
20     Status            25128 non-null    int64
```



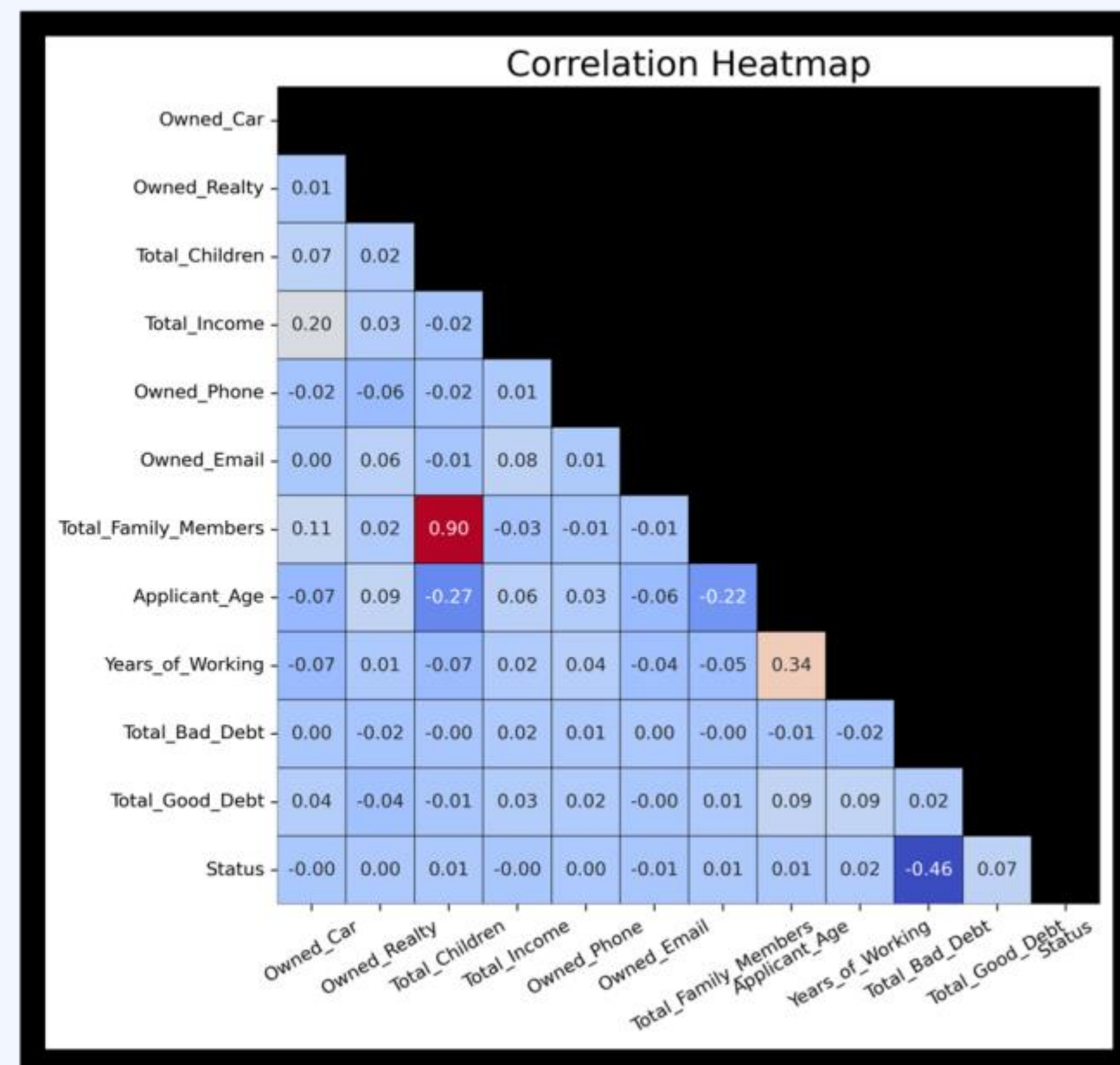


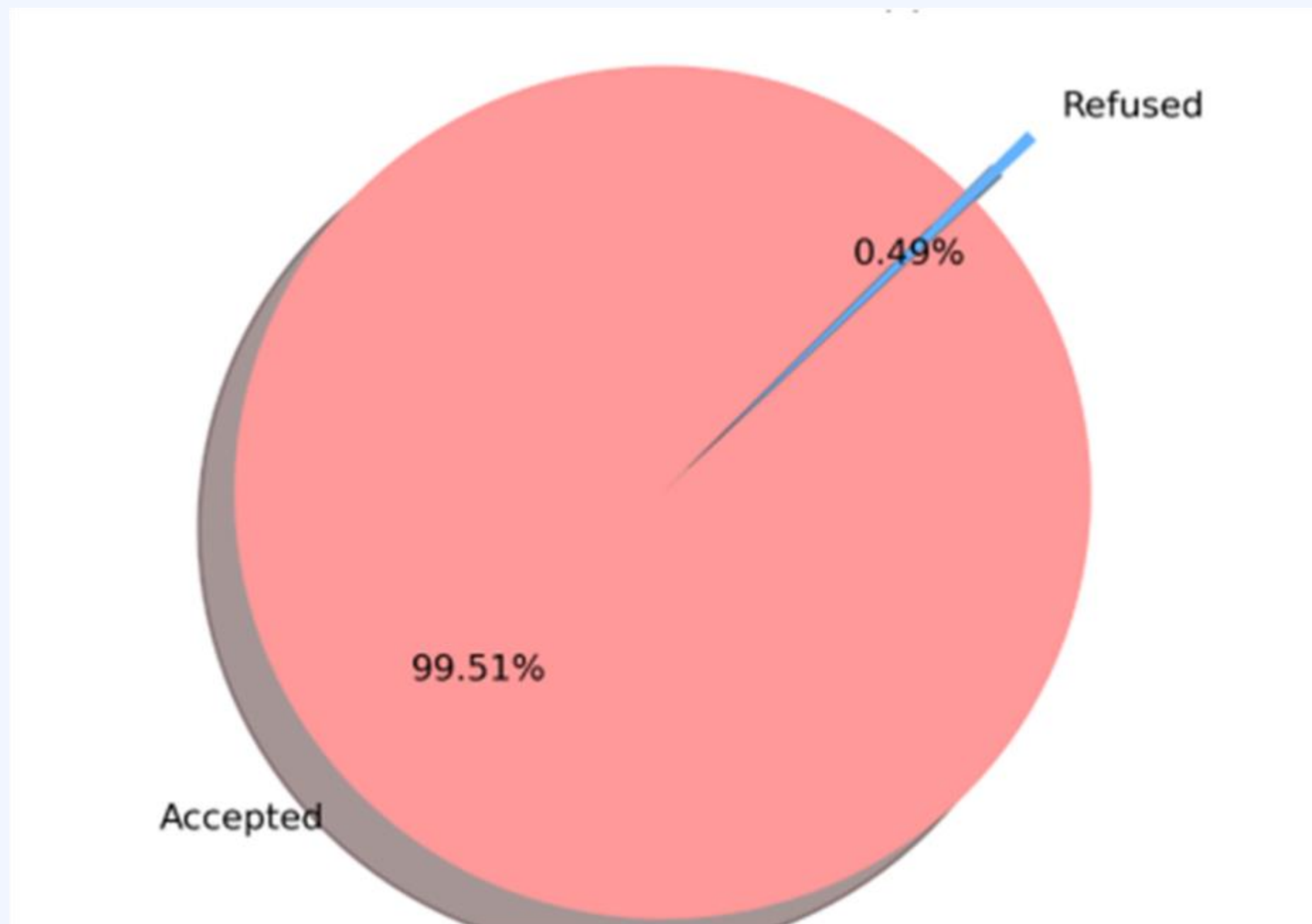
# 기술 통계 & HEATMAP

Presentation for Data Visualization



	count	mean	std	min	25%	50%	75%	max
Applicant_ID	25128.00	5078835.48	41943.78	5008806.00	5042225.75	5079004.00	5115603.25	5150487.00
Owned_Car	25128.00	0.42	0.49	0.00	0.00	0.00	1.00	1.00
Owned_Realty	25128.00	0.65	0.48	0.00	0.00	1.00	1.00	1.00
Total_Children	25128.00	0.51	0.76	0.00	0.00	0.00	1.00	5.00
Total_Income	25128.00	194836.50	104521.12	27000.00	135000.00	180000.00	225000.00	1575000.00
Owned_Mobile_Phone	25128.00	1.00	0.00	1.00	1.00	1.00	1.00	1.00
Owned_Work_Phone	25128.00	0.27	0.45	0.00	0.00	0.00	1.00	1.00
Owned_Phone	25128.00	0.29	0.46	0.00	0.00	0.00	1.00	1.00
Owned_Email	25128.00	0.10	0.30	0.00	0.00	0.00	0.00	1.00
Total_Family_Members	25128.00	2.29	0.93	1.00	2.00	2.00	3.00	7.00
Applicant_Age	25128.00	41.00	9.55	21.00	33.00	40.00	48.00	68.00
Years_of_Working	25128.00	7.69	6.42	1.00	3.00	6.00	10.00	44.00
Total_Bad_Debt	25128.00	0.33	1.57	0.00	0.00	0.00	0.00	49.00
Total_Good_Debt	25128.00	21.06	14.74	1.00	9.00	18.00	31.00	61.00
Status	25128.00	1.00	0.07	0.00	1.00	1.00	1.00	1.00





데이터 불균형 존재

- 승인 99.51%

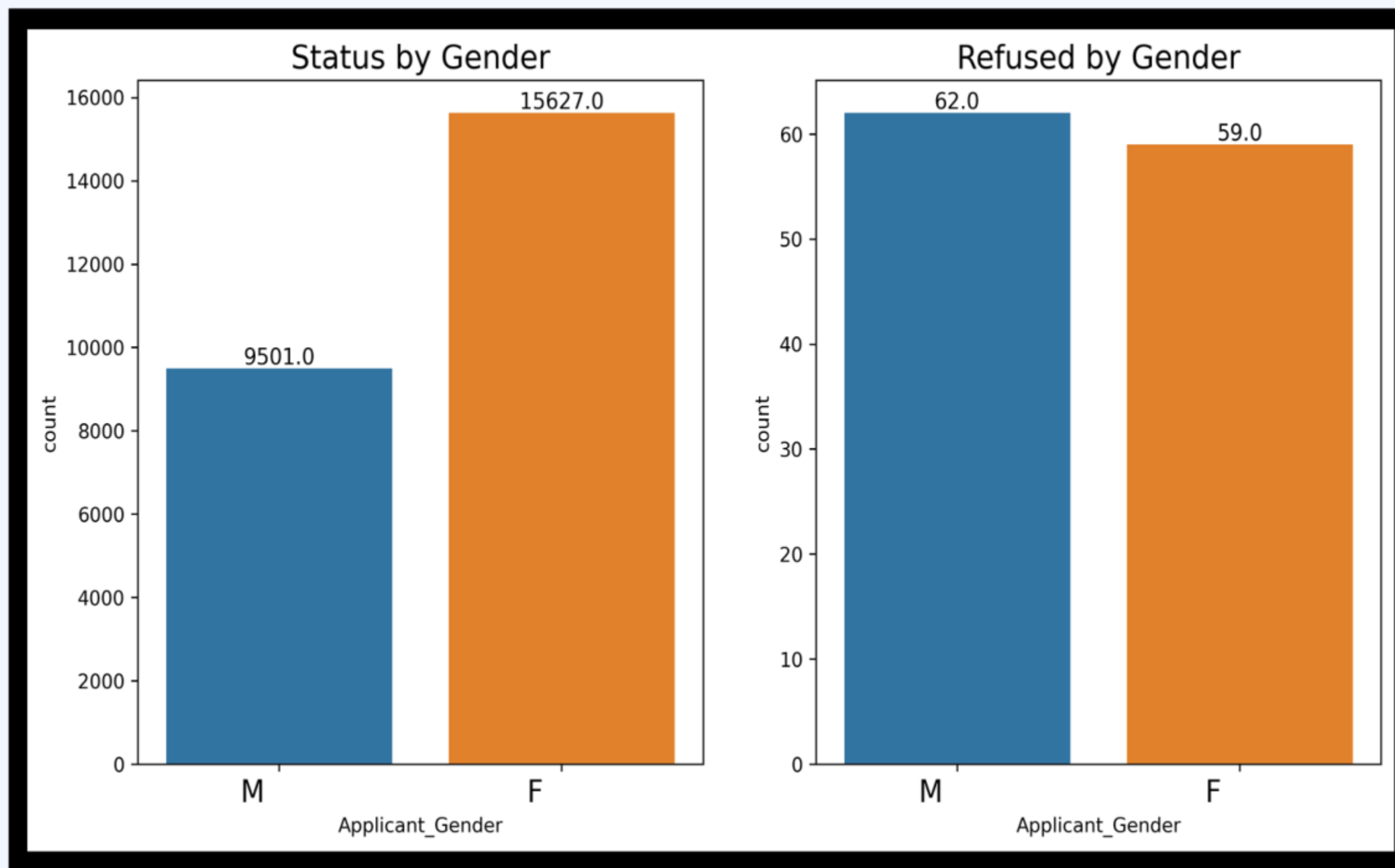
- 거절 0.49%

데이터 재구성

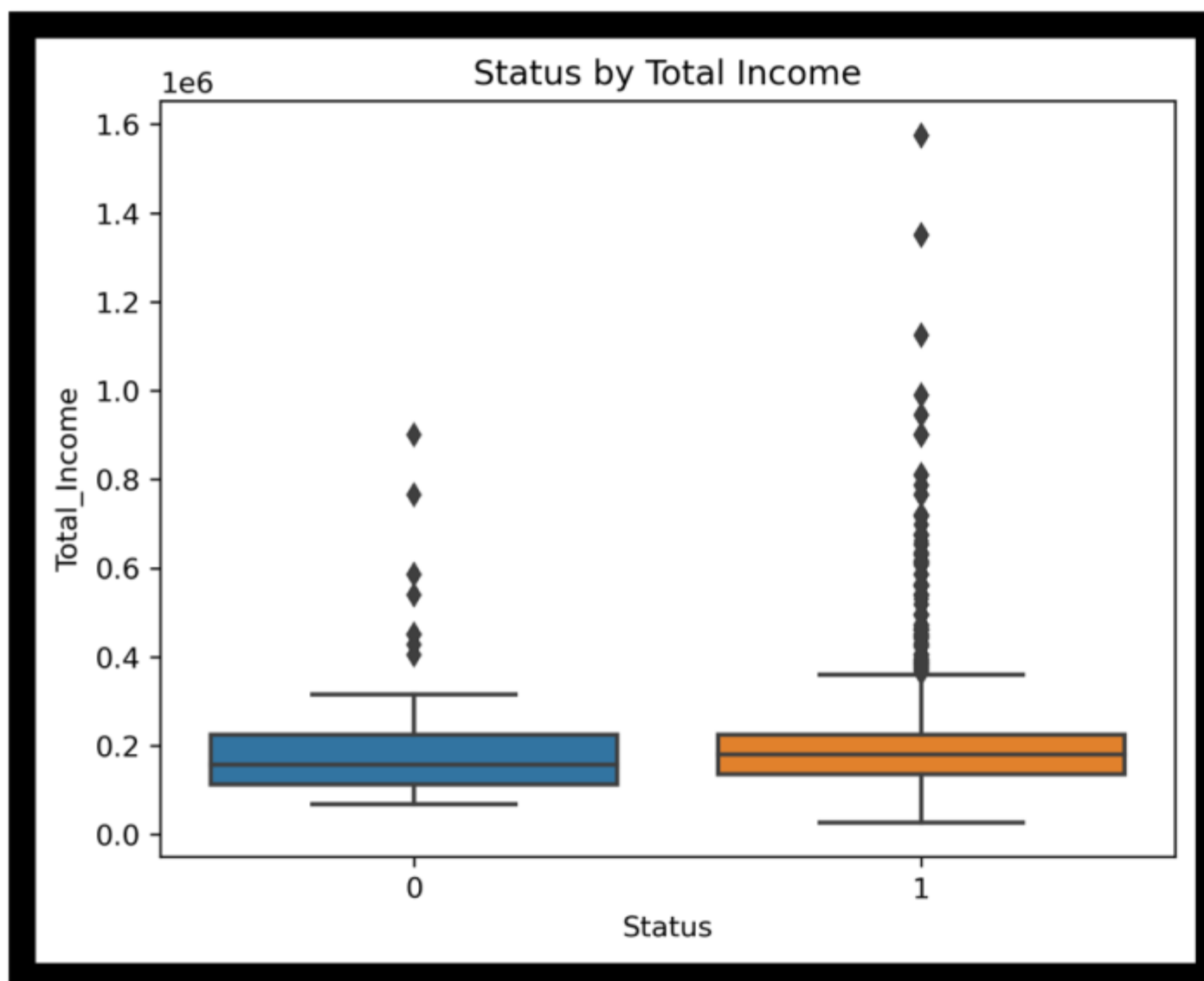
- Filtering

- 오버 & 언더 샘플링

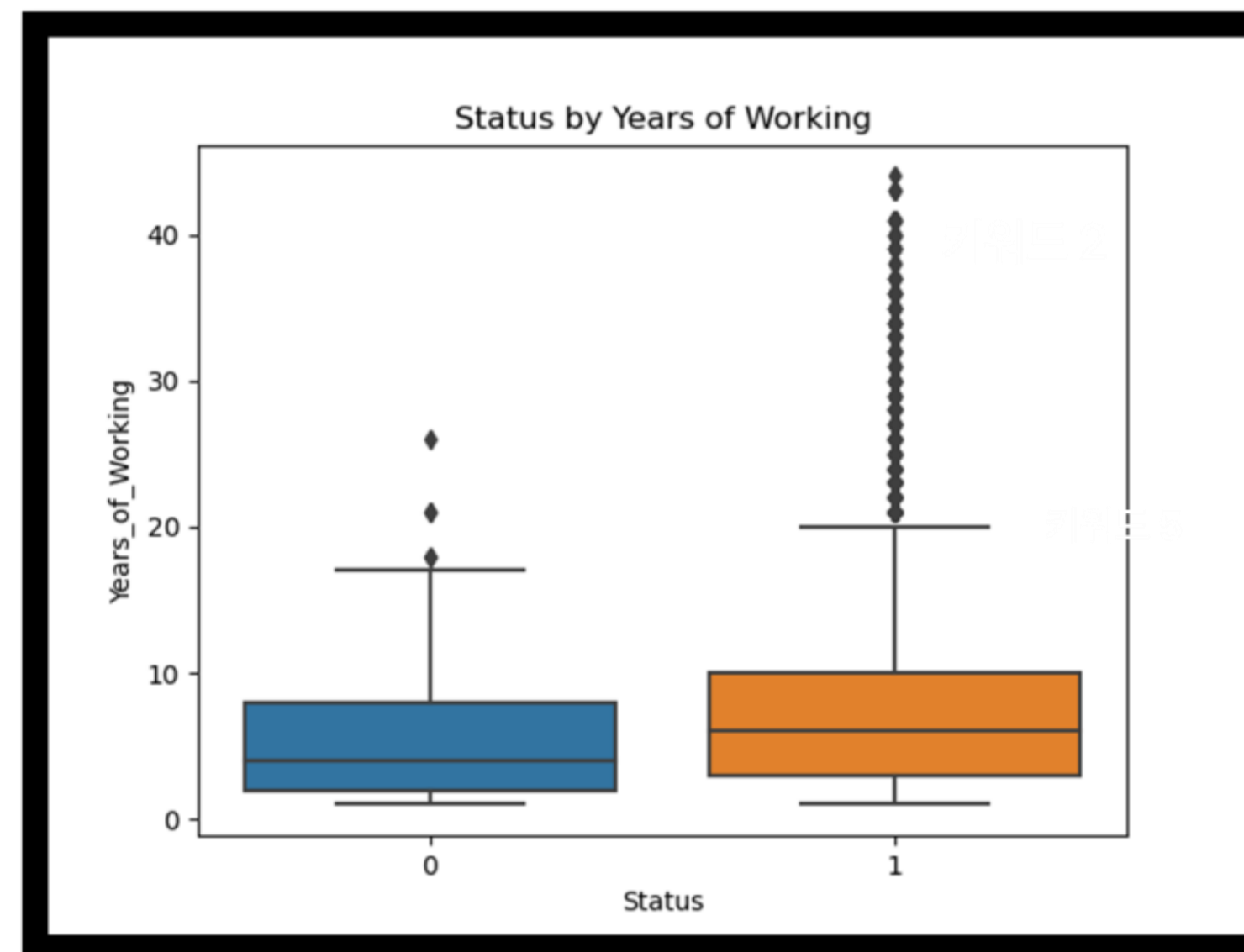
# 성별에 따른 신용카드 승인 현황



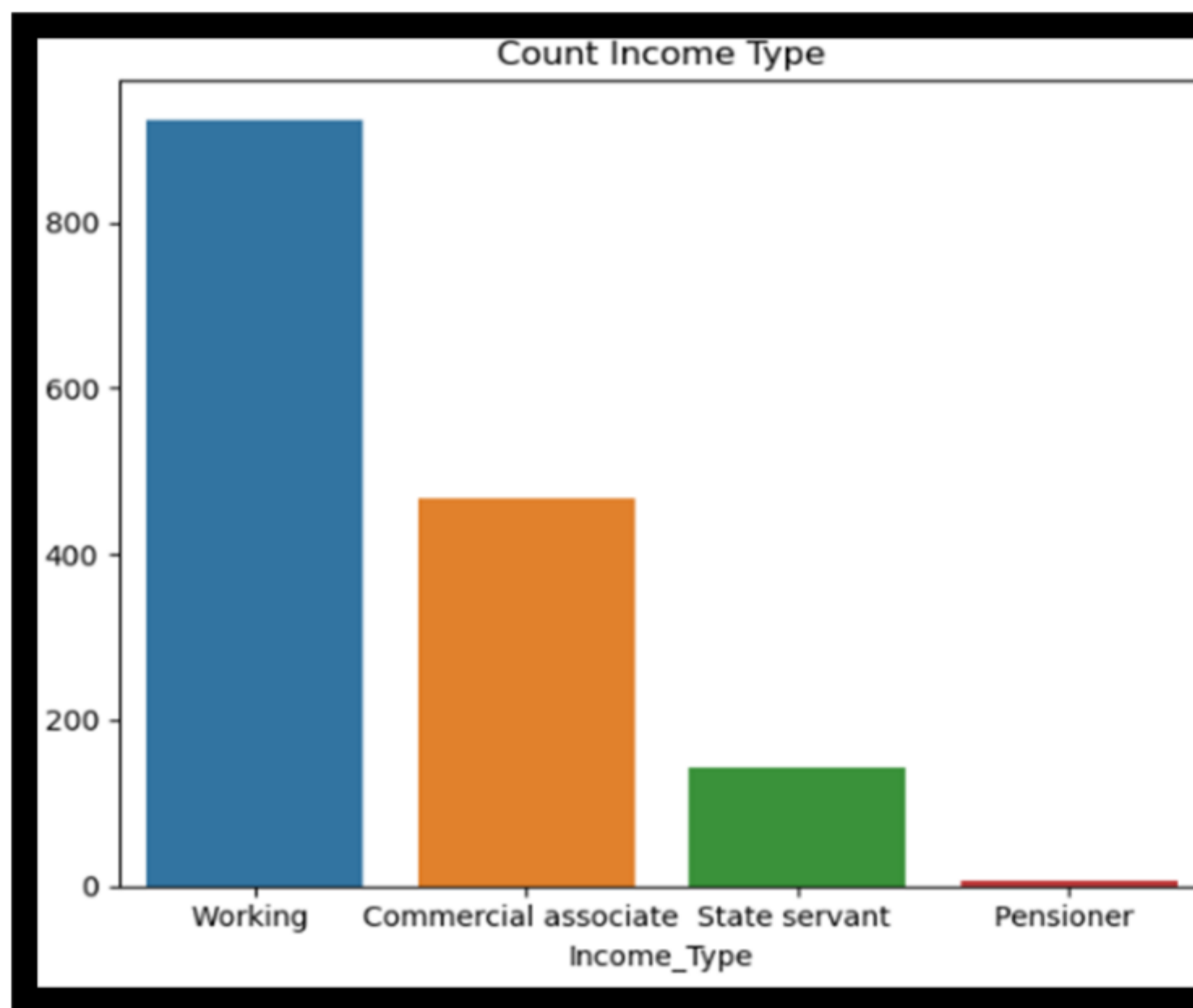
## 총수입별 현황



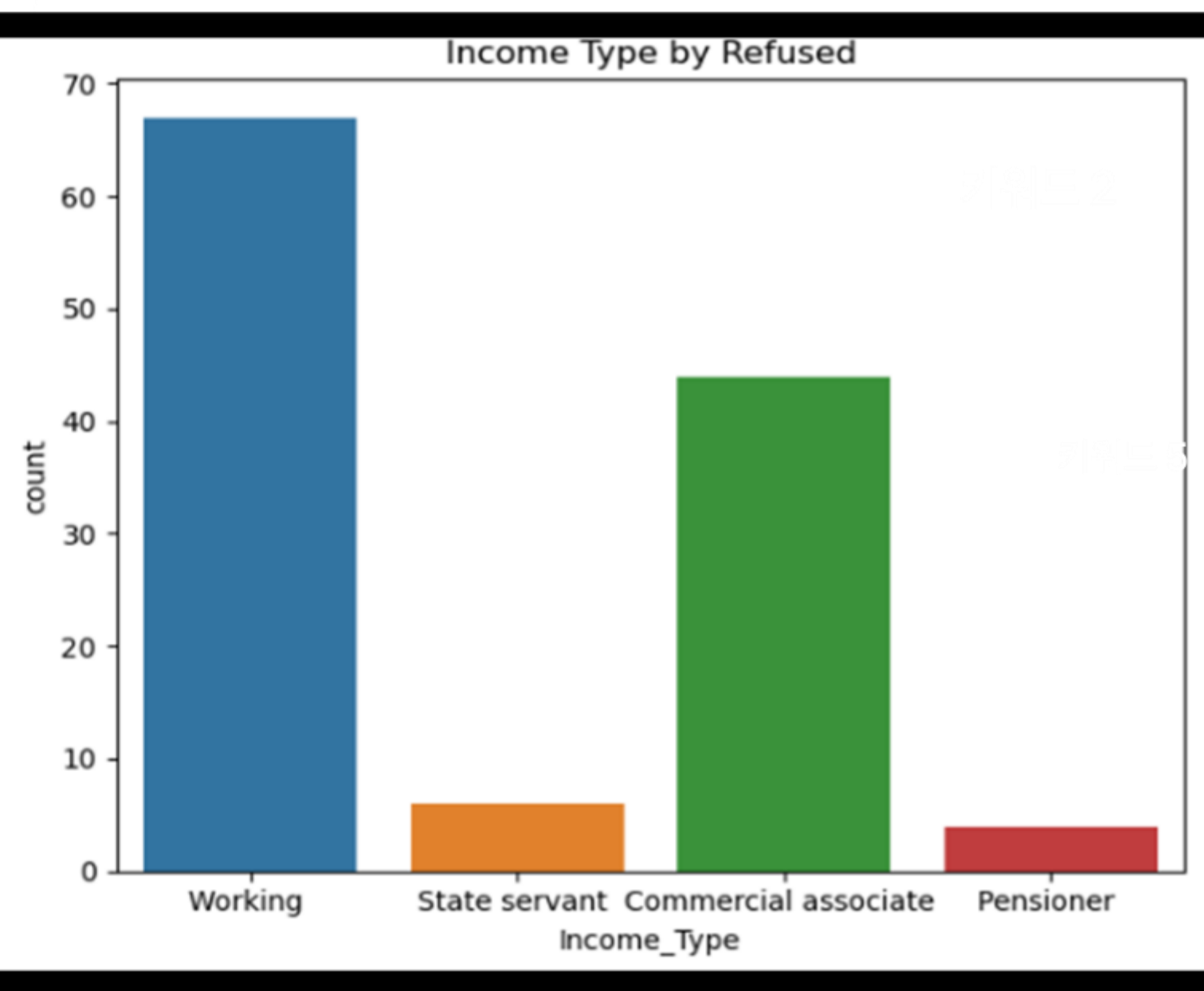
## 근무연수 별 현황



## 직업별 분포

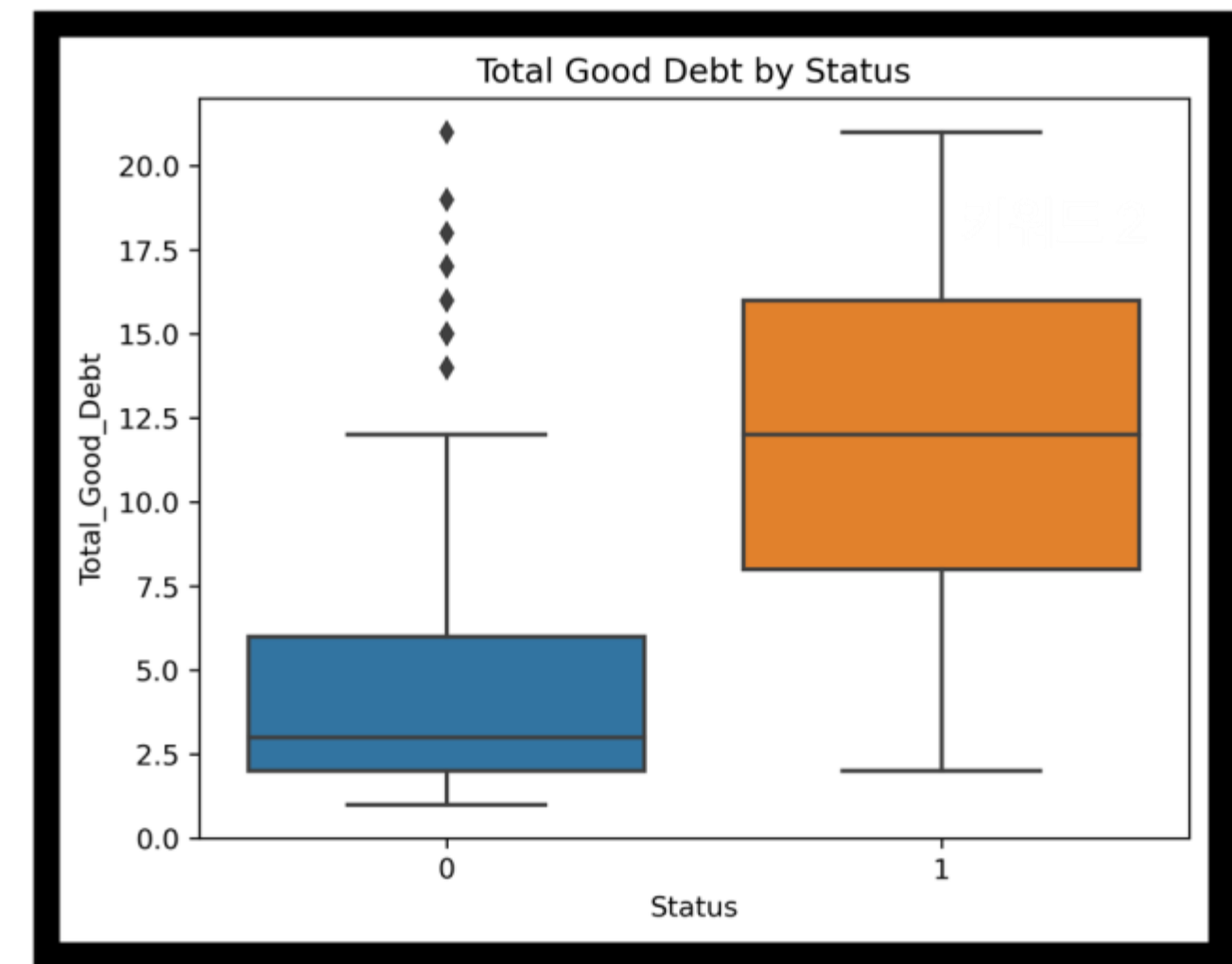
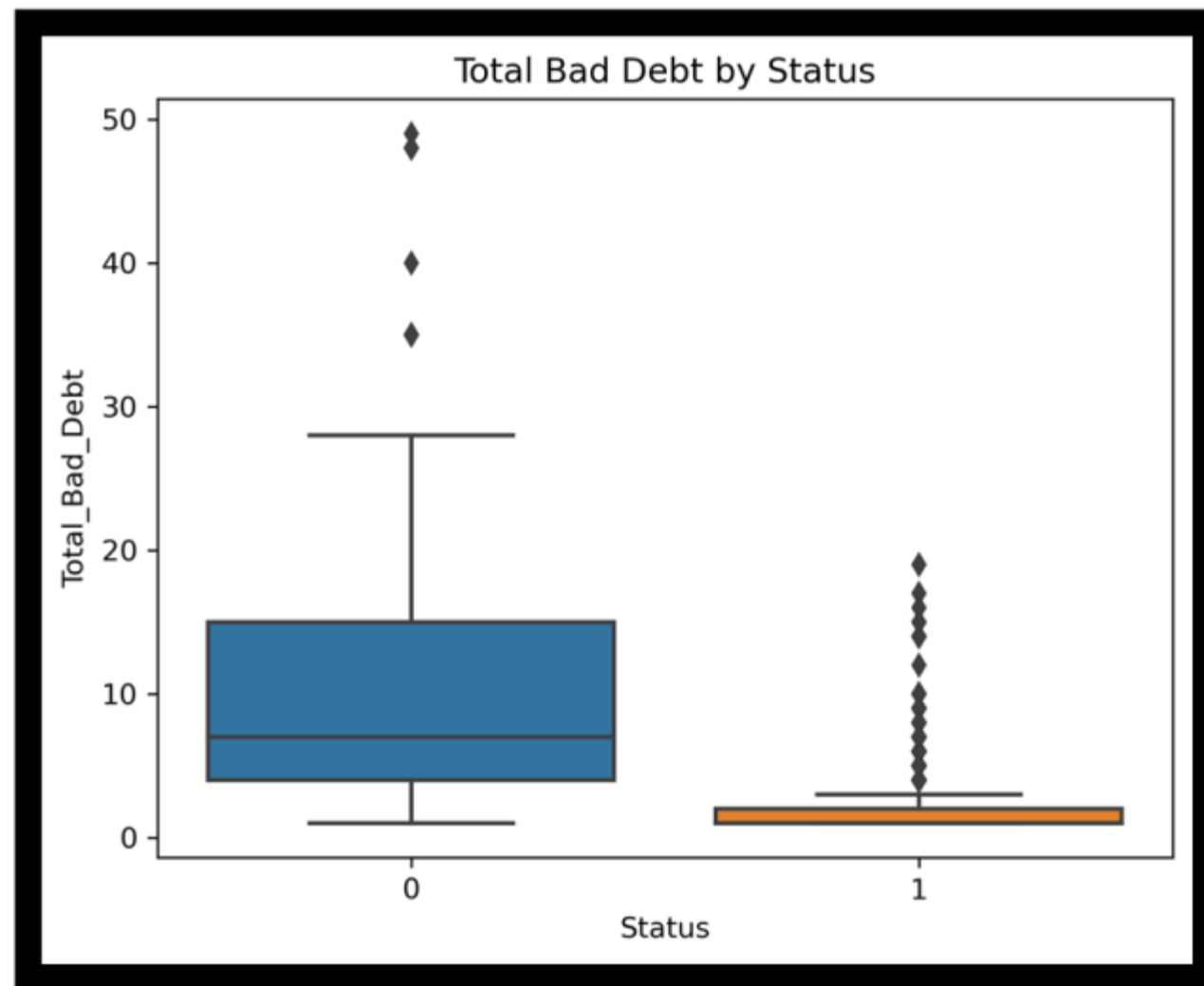


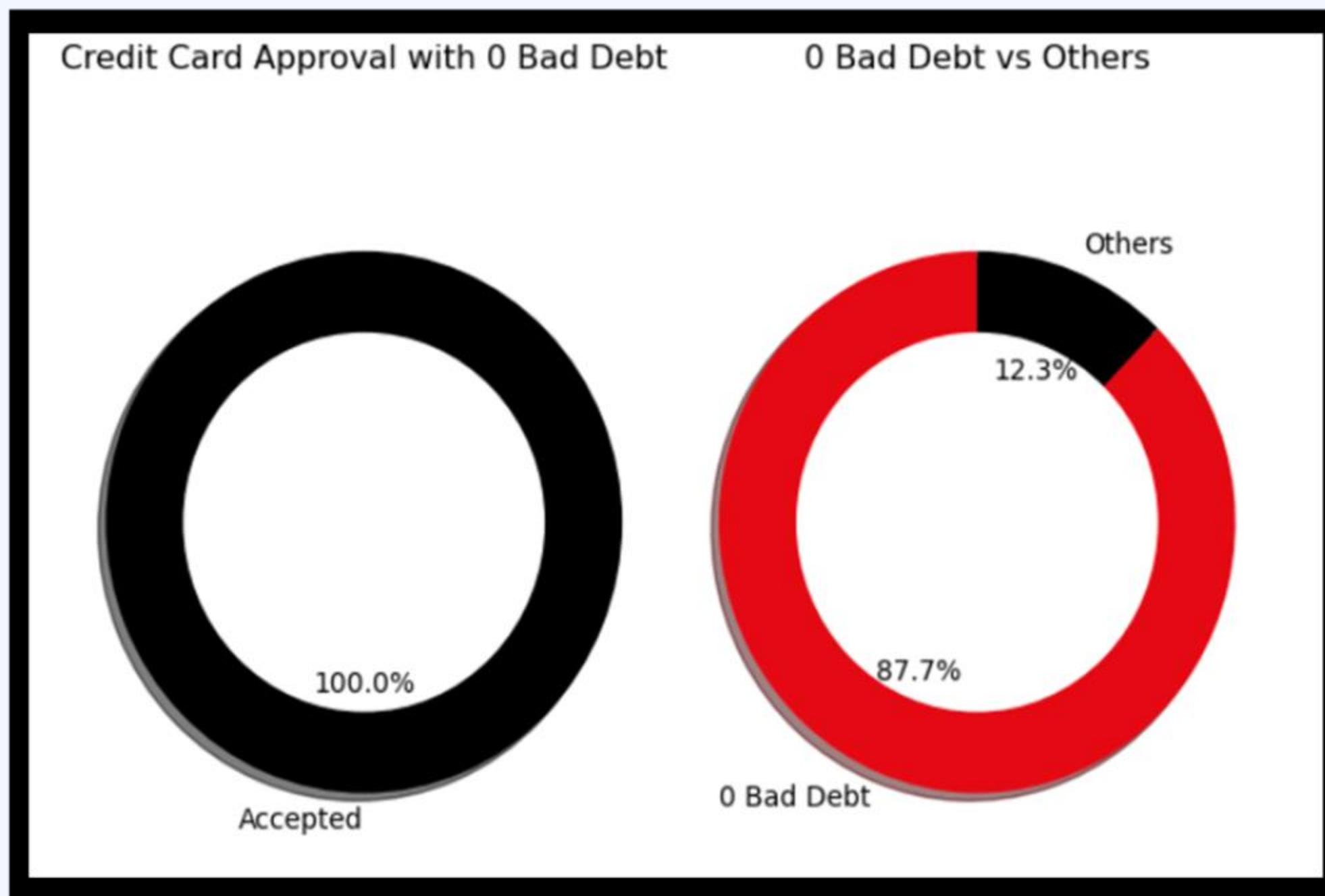
## 직업별 거절 분포





## 부실채권

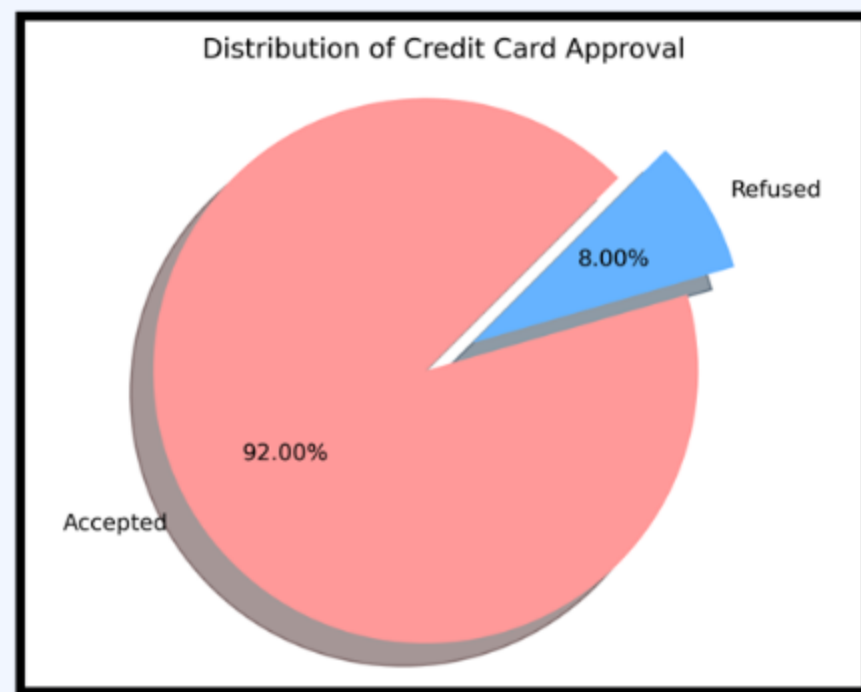




"Totla bad Debt"가 0인 기록 제거

데이터 불균형 감소

모델 소요 시간 축소



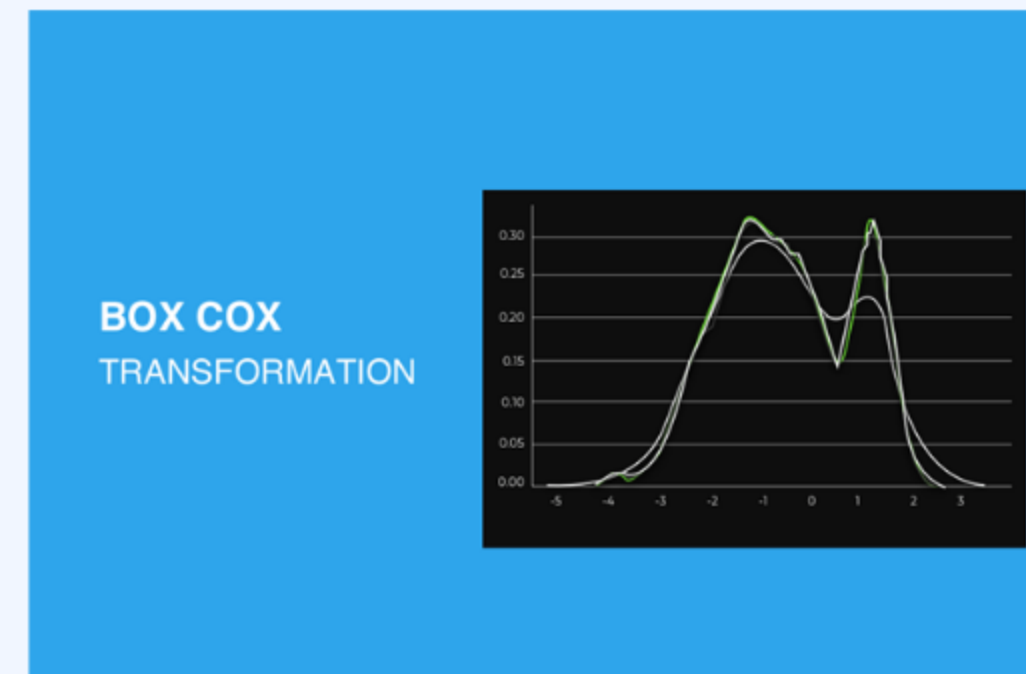
## 변수 엔지니어링

SMOTE 방법을 통한 재 샘플링  
균형 잡힌 데이터 세트를 생성



## 가설 검증 및 특징 제거

T 검정  
5개 변수 제거



## BOX-COX

왜도와 첨도를 Box-Cox 변환  
비대칭성 해결  
무거운 꼬리 감소

# 모델링 및 평가

1

## 모델 선택

다양한 머신러닝 알고리즘 중 가장 적절한 모델을 선택하여 예측 모델링을 수행합니다.

2

## 학습 및 테스트

데이터를 학습과 테스트 세트로 나눠 모델을 학습하고 예측 성능을 평가합니다.

3

## 모델 성능 향상

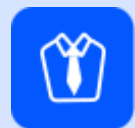
하이퍼파라미터 튜닝 및 앙상블 기법을 활용하여 모델의 성능을 최대화합니다.





# 모델 선택

Presentation for Data Visualization



Model 01

ADABOOST

오분류 데이터에 집중 학습하는 앙상블 기법



Model 02

XGBOOST

경사 하강법, 자동 조정, 트리 구조 등을 도입한 앙상블 기법



Model 03

LOGISTIC  
REGRESSION

로지스틱 함수를 사용하여 데이터를 분류하는 통계 기법



```

AdaBoost Classification Report:
              precision    recall  f1-score   support

         0       0.97       0.92       0.94         121
         1       1.00       1.00       1.00        25007

 accuracy          1.00        25128
 macro avg          0.98        25128
 weighted avg       1.00        25128

XGBoost Classification Report:
              precision    recall  f1-score   support

         0       0.97       0.93       0.95         121
         1       1.00       1.00       1.00        25007

 accuracy          1.00        25128
 macro avg          0.99        25128
 weighted avg       1.00        25128

Logistic Regression Classification Report:
              precision    recall  f1-score   support

         0       0.95       0.58       0.72         121
         1       1.00       1.00       1.00        25007

 accuracy          1.00        25128
 macro avg          0.97        25128
 weighted avg       1.00        25128

```

## 교차 검증

- 5개의 분할을 가진 StratifiedKFold 교차 검증 전략

## SMOTE

- 소수 클래스에 대한 합성 샘플을 생성하여 균형잡힌 분포 생성
- 모델이 클래스의 중요성을 더 정확히 반영

## 주성분 분석

- 변동성이 큰 주성분에 초점을 맞추어 필수 정보를 유지하면서 중복된 특성을 제거

## 기타 기술

- 필터링과 같은 모델 성능을 향상시키기 위한 기술 적용



# 모델 성능 향상

Presentation for Data Visualization

## 하이퍼 파라미터 조정

### 서브 샘플

각 트리가 데이터 샘플의 100%(1.0)로 훈련되도록 보장

### 정규화 람다

과적합을 피하기 위해 가중치에 L2 정규화 항을 1로 적용

### 정규화 알파

특징 선택을 위해 가중치에 0.1 값의 L1 정규화 항을 통합

### Number of estimator

앙상블에서 900개의 트리를 사용

### Min\_child\_weight

자식에 필요한 인스턴스 가중치의 합을 3으로 설정

### Max depth

각 트리가 최대 8개의 가지를 생성하도록 설정



5개 결과 모두 "거절됨"을 잘 예측하는 모델



모든 클래스에 높은 정확도를 유지하는  
균형잡힌 모델



신용카드 승인 절차를 자동화할 수  
있는 신뢰 있는 모델

Fold 1 Classification Report:				
	precision	recall	f1-score	support
0	0.96	1.00	0.98	25
1	1.00	1.00	1.00	284
accuracy			1.00	309
macro avg	0.98	1.00	0.99	309
weighted avg	1.00	1.00	1.00	309
Fold 2 Classification Report:				
	precision	recall	f1-score	support
0	0.96	1.00	0.98	24
1	1.00	1.00	1.00	284
accuracy			1.00	308
macro avg	0.98	1.00	0.99	308
weighted avg	1.00	1.00	1.00	308
Fold 3 Classification Report:				
	precision	recall	f1-score	support
0	0.89	1.00	0.94	24
1	1.00	0.99	0.99	284
accuracy			0.99	308
macro avg	0.94	0.99	0.97	308
weighted avg	0.99	0.99	0.99	308
Fold 4 Classification Report:				
	precision	recall	f1-score	support
0	0.92	1.00	0.96	24
1	1.00	0.99	1.00	284
accuracy			0.99	308
macro avg	0.96	1.00	0.98	308
weighted avg	0.99	0.99	0.99	308
Fold 5 Classification Report:				
	precision	recall	f1-score	support
0	0.96	1.00	0.98	24
1	1.00	1.00	1.00	284
accuracy			1.00	308
macro avg	0.98	1.00	0.99	308
weighted avg	1.00	1.00	1.00	308





## IMPORTANT FEATURE

Rank	Pearson's Correlation	L1 Regularization	XGBoost
1	Total_Bad_Debt	P.Total_Bad_Debt	Bad_Debt_to_Good_Debt_Ratio
2	Bad_Debt_to_Good_Debt_Ratio	P.Total_Good_Debt	Total_Good_Debt
3	P.Total_Bad_Debt	Total_Good_Debt	Income_Type_3
4	Total_Good_Debt	Total_Bad_Debt	Income_Type_2
5	Total_Good_Debt	Bad_Debt_to_Good_Debt_Ratio	Income_Type_1
6	Age_to_Working_Years_Ratio	Housing_Type	Education_Type_1



피어슨 상관 계수



L1 정규화

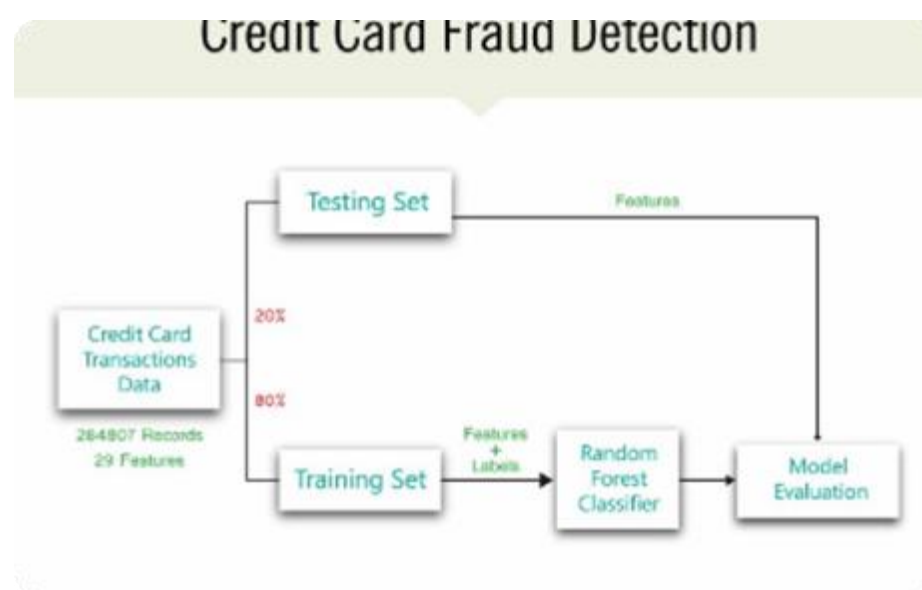


앙상블



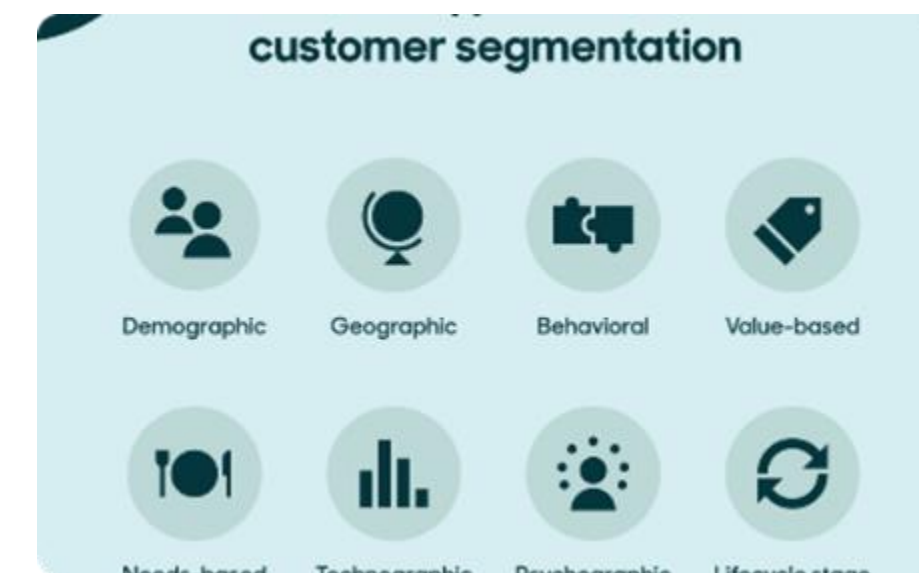
## 은행

- 은행에서 신용카드 승인 예측을 활용하여 대출 신청에 대한 신속한 판단을 도와줍니다.



## 금융 회사

- 금융 회사에서는 신용카드 사기 탐지를 위해 승인 예측 모델을 사용합니다.



## 커머스

- 커머스 기업에서는 고객 세분화를 위해 신용카드 승인 예측을 활용합니다.



## 한계



## 고객 데이터 의존

고객 정보만을 포함하는 데이터셋에 의존

## 결론



## 성공적인 예측

머신러닝 학습으로 신용카드 승인  
예측 모델을 성공적으로 구축