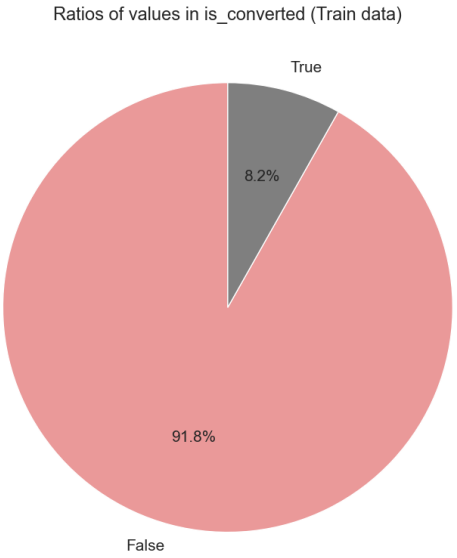
visualize_function

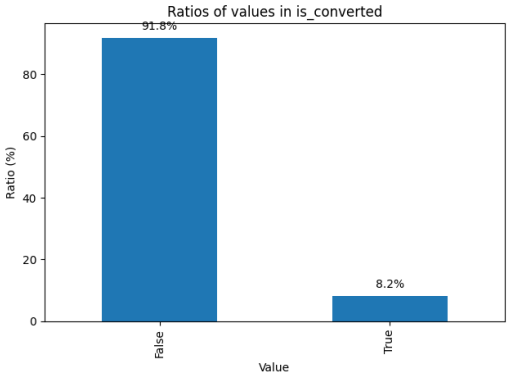
1. `def pie_value_ratios( df , column_name ):`

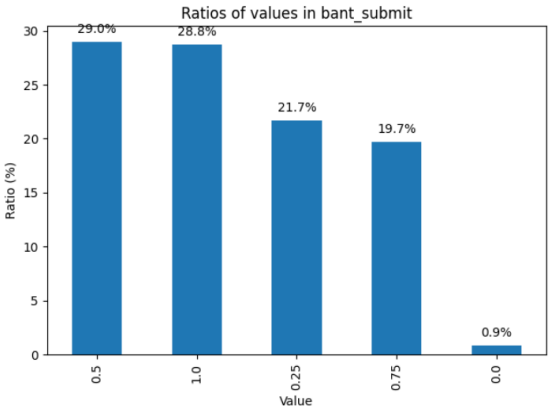
   각 feature 내의 True vs False 비율 파이차트 그리기



2. `def bar_value_ratios( df , column_name ):`

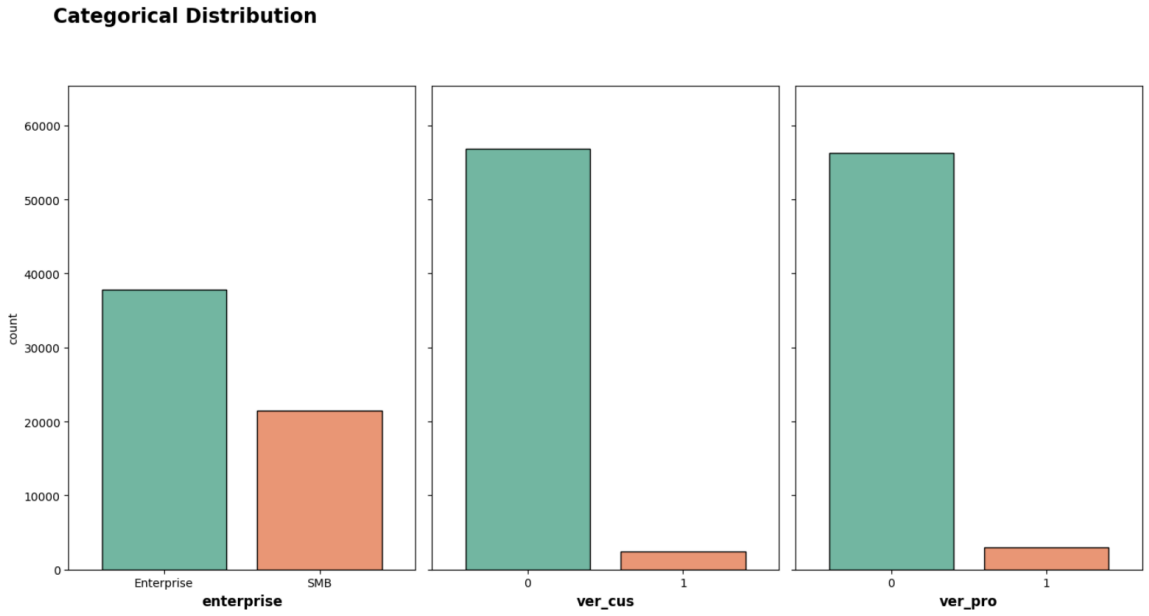
   각 feature 내의 class 분포 막대그래프





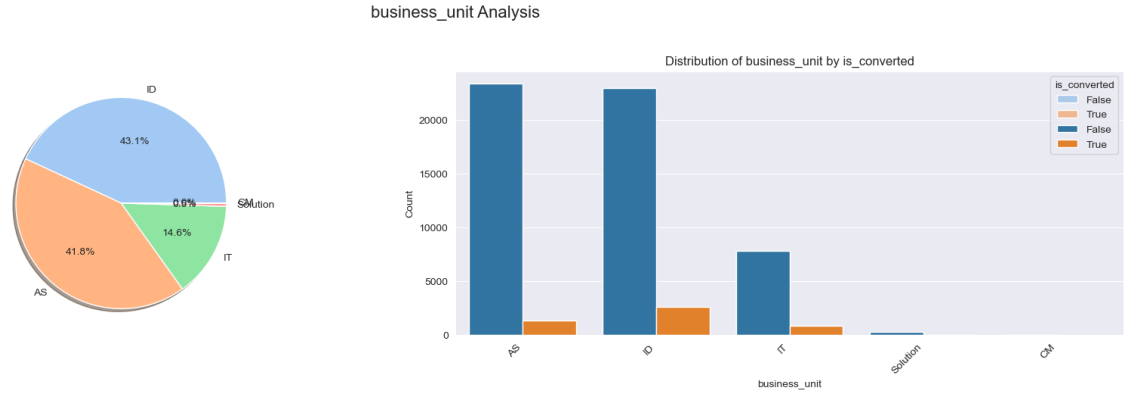
3. `def plot_categorical_distribution(df, column_names ):`

   feature 내 class count 한번에 비교(true ratio X)



4. `def pie_bar_cat_column( df_input , column_name , target ):`

   각 (cateogory) featrue 내 class 분포

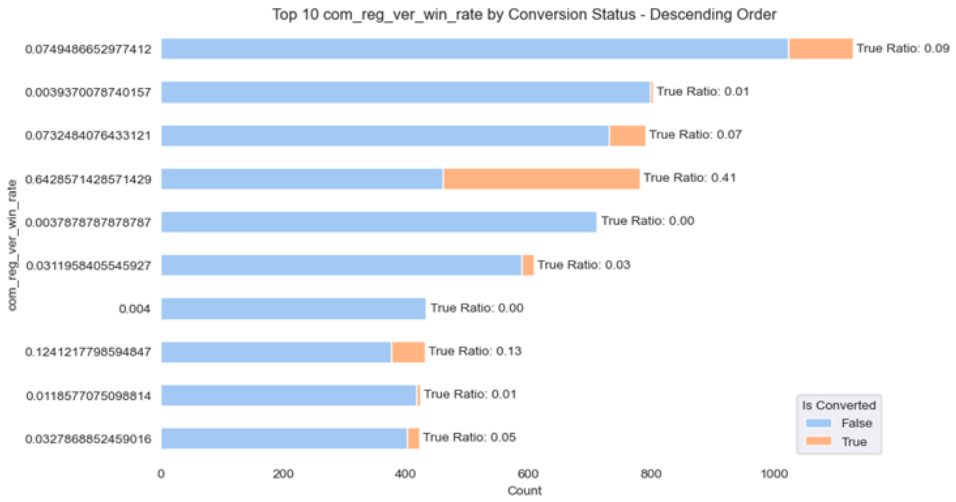


5. `def bar_top_categories_with_ratio( df , column_name , top_n =10, target ='is_converted'):`

   {data_count 기준 내림차순으로 상위 10개 시각화 + True_ratio}

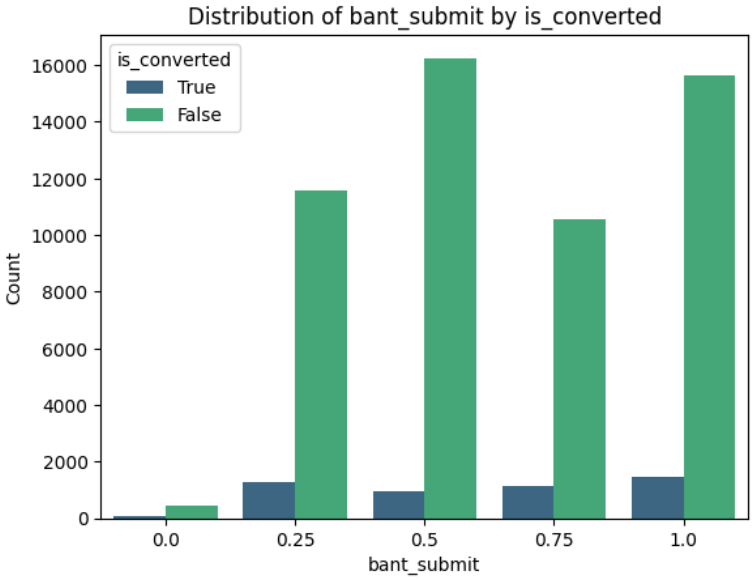
   - 변수 내 class 10개 이상인 경우, class_count 갯수 내림차순으로 상위 10개만 뽑아 시각화

   - True ratio : 각 class내에 Ture/(False+Ture)



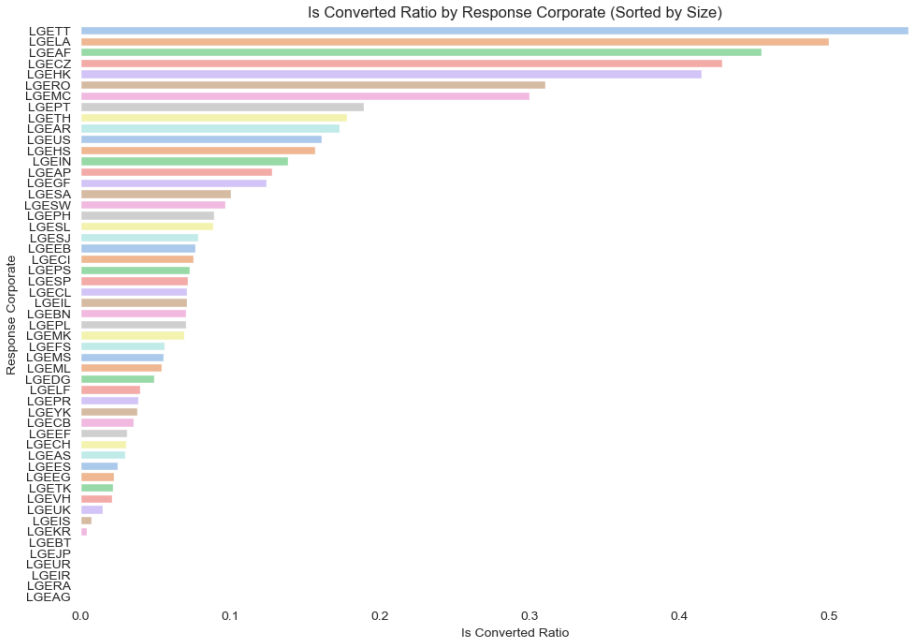
6. `def double_bar_cat_columns( df_input , categorical_vars , target ):`

   각 feature 내 클래스의 False, True bar 따로 그림



7. `def bar_true_ratio( df_input , column_name ):`

   클래스별 is_converted 비율만 보기

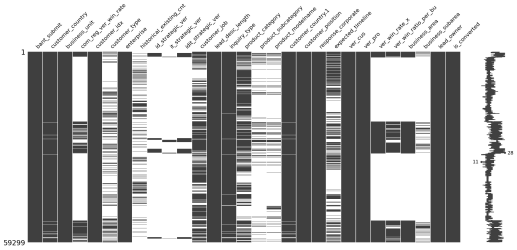


8. `def show_binary_ratio( df_input , column_name : str , target : str ):`

   각 feature 내 class의 true ratio

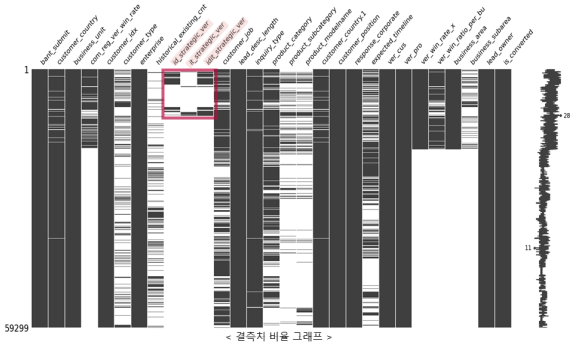
```
business_unit: 59299
AS        True Ratio: 0.06, Total Count: 24774
CM        True Ratio: 0.00, Total Count: 2
ID        True Ratio: 0.10, Total Count: 25563
IT        True Ratio: 0.10, Total Count: 8664
Solution  True Ratio: 0.01, Total Count: 296
```

9. `msno.matrix(df_train)` (결측치 시각)



`msno.matrix(sorted_bus_area)`

(특정 컬럼 기준 정렬된 결측치 시각)

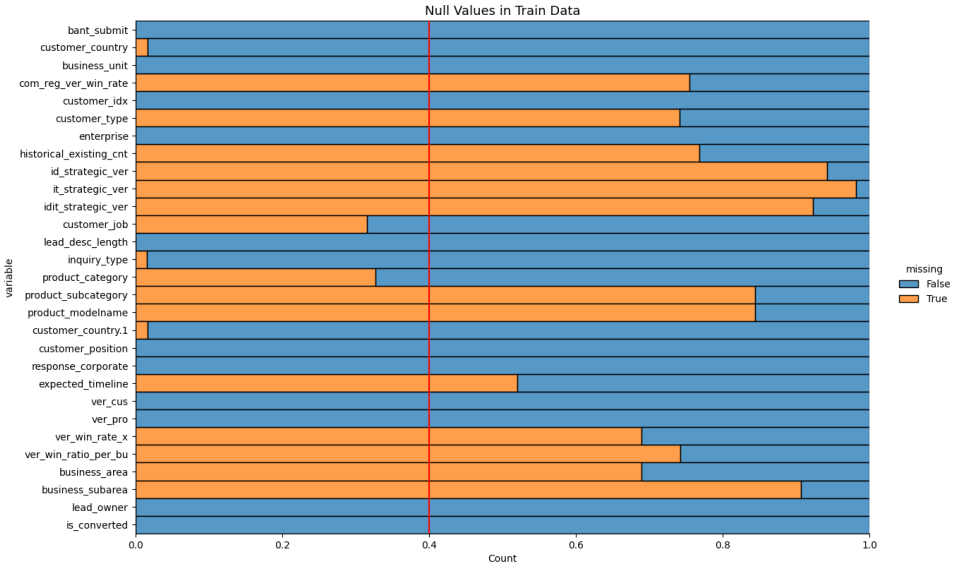


10. `def show_missing_rates( df_input , selected_columns =None):` (결측치 비율)

```
business_unit      :  0.00% missing values (object)
id_strategic_ver   : 94.19% missing values (float64)
it_strategic_ver   : 98.11% missing values (float64)
```

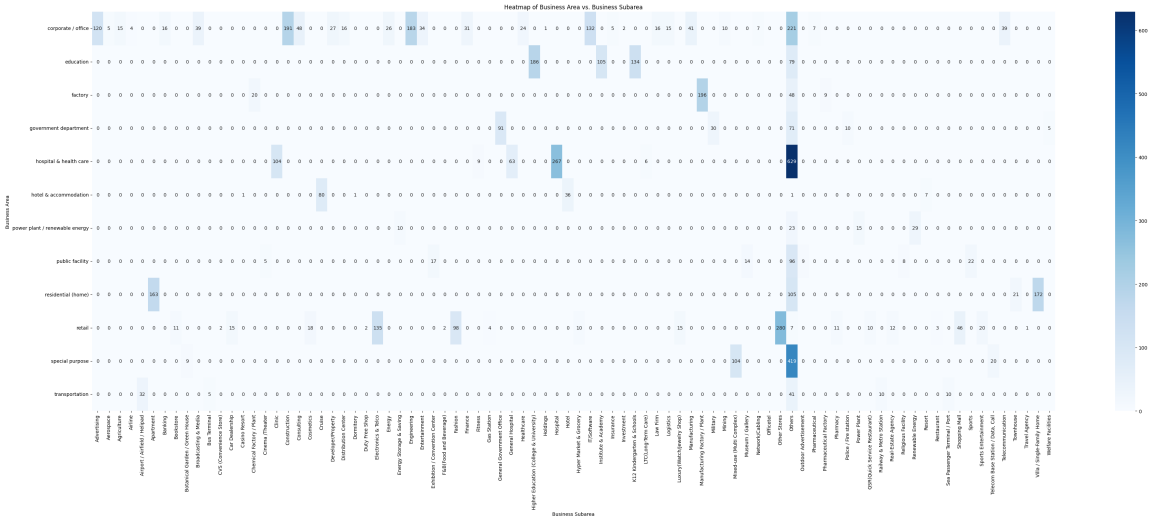
11. `def plot_missing( df_input ):`

결측치 분포



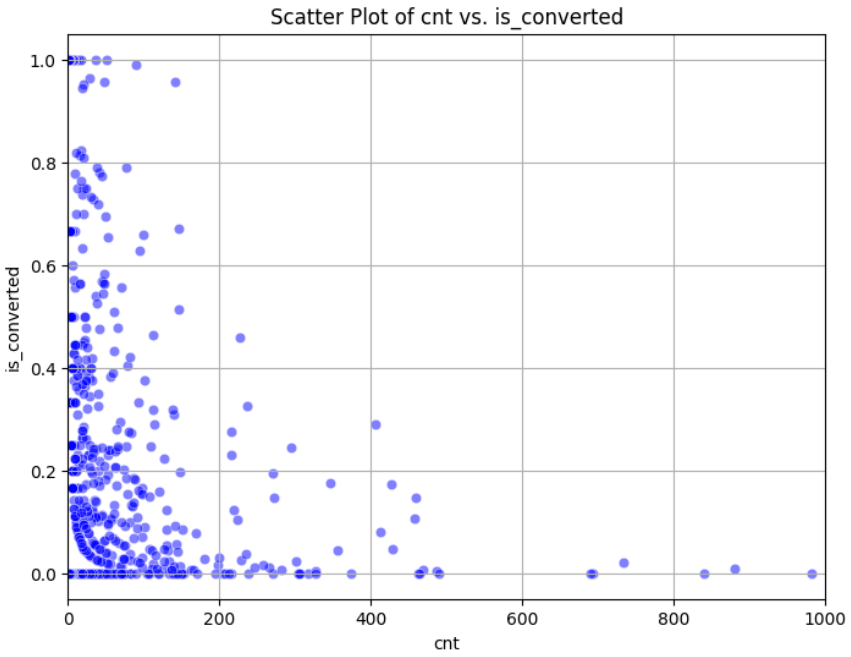
12. `def compare_features_class( df_input , cloumn_names : list ):`

두 feature 간 class 연관성 비교 → 일대일 대응인지 다대일 대응인지 관계 알 수 있음



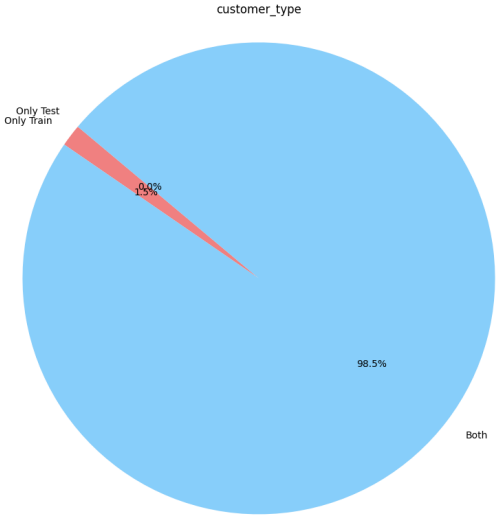
13. `def scatter_column_ratio( df_input , colume_name : str ):`

특정 feature 내 True 분포를 산점도로 나타내기



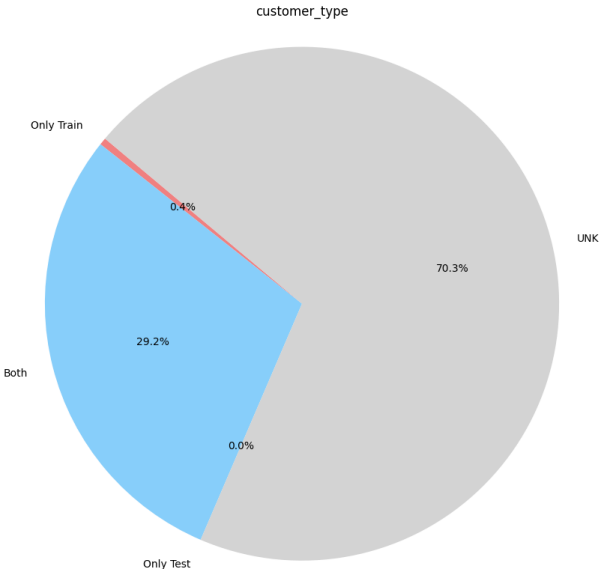
14. `def plot_pie_chart( train , test , feat_names ):`

    train, test 존재하는 데이터 분포 알기



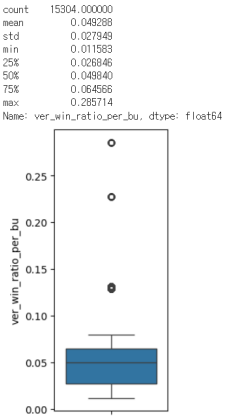
15. `def plot_pie_chart_fillna( train , test , feat_names ):`

    train, test 존재하는 데이터 분포 알기



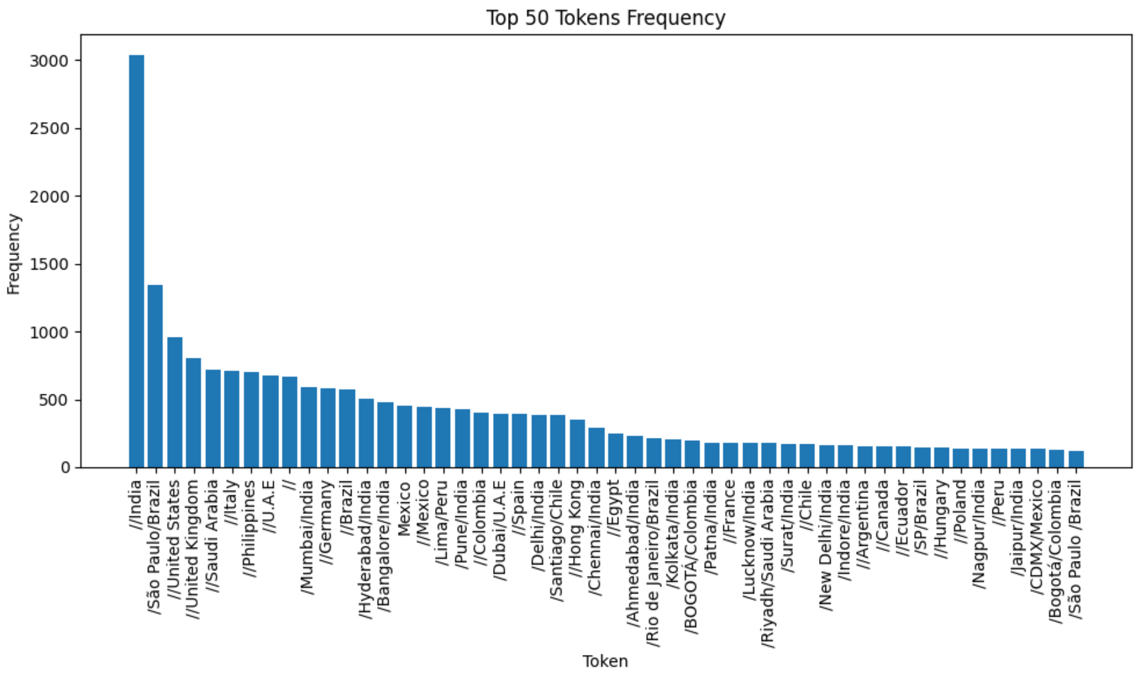
16. `def describe_boxplot( df_input , column_name ):`

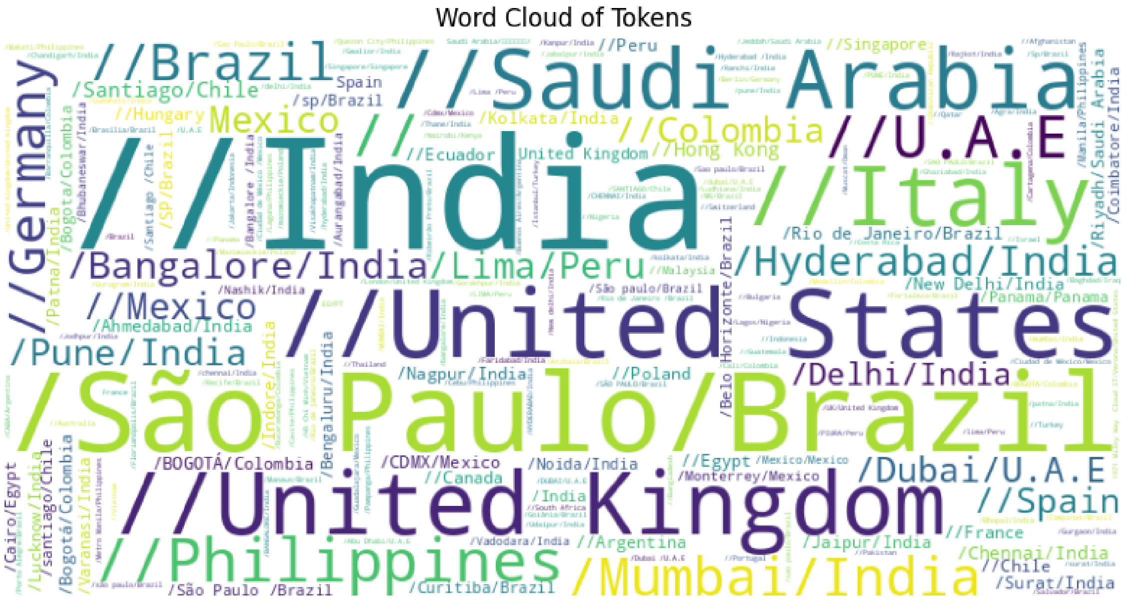
    featrue describe boxplot



17. `def wordcloud_visualize( df ):`

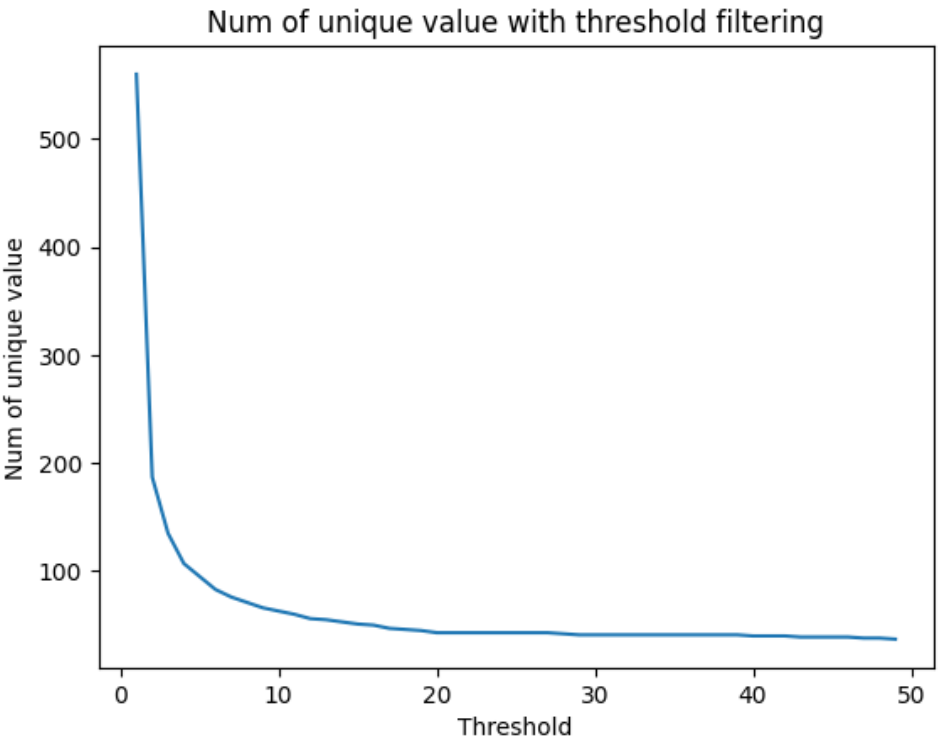
    feature 내 상위 50개 클래스 분포 막대그래프 & 워드 클라우드 시각화





18. `def threshold_counts( df_input , column_name ):`

feature 내 n 이상 , m 미만의 개수 이상 나오는 것들만 필터링


Num of unique value with threshold filtering
Num of unique value
500
400
300
200
100
0
10
20
30
40
50
Threshold