Objective: To explore various AutoEDA capabilities and perform analysis on a given dataset

This notebook will focus on DataPrep

2. AutoEDA - DataPrep

Dataset Reference: Loan Prediction dataset from Kaggle

Features:

- General Overview - Quick insights of all variables in the dataset using the plot dataframe.
- Details about each variables / features in the dataset by using create_report - overview, variables, interactions, correlations, missing values
- Interactions - based on x-axis and y-axis scatter plots
- Correlations between variables - Pearson's Correlation Coefficient, Spearman's Rank Correlation Coefficient, Kendall's Rank Correlation Coefficient
- Missing Values - Bar chart, Spectrum, Heatmap, Dendogram representations
- We can pick one particular feature and analyze - Stats, Bar chart, Pie chart, Word Count, Word Frequency etc as per applicability

When To Use?

- Dataset size is fairly very large (this seems to be 10X faster than Pandas Profiling tools due to it's highly optimized Dask-based computing module)
- Need some quick insights about an unknown dataset
- Use this as a basis for your further EDA analysis on top of it

```python
import pandas as pd

df_train = pd.read_csv("../input/loan-eligible-dataset/loan-train.csv")

df_train.head()
```

Out[1]:

Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_Hist

	Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_Hist
0	LP001002	Male	No	0	Graduate	No	5849	0.0	NaN	360.0	
1	LP001003	Male	Yes	1	Graduate	No	4583	1508.0	128.0	360.0	
2	LP001005	Male	Yes	0	Graduate	Yes	3000	0.0	66.0	360.0	
3	LP001006	Male	Yes	0	Not Graduate	No	2583	2358.0	120.0	360.0	
4	LP001008	Male	No	0	Graduate	No	6000	0.0	141.0	360.0	

```
In [2]:  df_test = pd.read_csv("../input/loan-eligible-dataset/loan-test.csv")

         df_test.head()
```

Out[2]:

	Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_Hist
0	LP001015	Male	Yes	0	Graduate	No	5720	0	110.0	360.0	
1	LP001022	Male	Yes	1	Graduate	No	3076	1500	126.0	360.0	
2	LP001031	Male	Yes	2	Graduate	No	5000	1800	208.0	360.0	
3	LP001035	Male	Yes	2	Graduate	No	2340	2546	100.0	360.0	
4	LP001051	Male	No	0	Not Graduate	No	3276	0	78.0	360.0	

```
In [3]:  df_train.shape
```

Out[3]: (614, 13)

```
In [4]:  df_test.shape
```

Out[4]: (367, 12)

```
In [5]:  !pip install dataprep  # Please use it for the first time if it is not installed in your environment
```

```python
from dataprep.eda import create_report, plot, plot_correlation, plot_missing
```

```
Collecting dataprep
  Downloading dataprep-0.3.0-py3-none-any.whl (1.8 MB)
     |████████████████████████████████| 1.8 MB 834 kB/s eta 0:00:01
Requirement already satisfied: numpy<2,>=1 in /opt/conda/lib/python3.7/site-packages (from dataprep) (1.19.5)
Requirement already satisfied: jinja2<3.0,>=2.11 in /opt/conda/lib/python3.7/site-packages (from dataprep) (2.11.3)
Requirement already satisfied: tqdm<5.0,>=4.48 in /opt/conda/lib/python3.7/site-packages (from dataprep) (4.59.0)
Collecting usaddress<0.6.0,>=0.5.10
  Downloading usaddress-0.5.10-py2.py3-none-any.whl (63 kB)
     |████████████████████████████████| 63 kB 1.2 MB/s eta 0:00:01
Requirement already satisfied: ipywidgets<8.0,>=7.5 in /opt/conda/lib/python3.7/site-packages (from dataprep) (7.6.3)
Requirement already satisfied: wordcloud<2.0,>=1.8 in /opt/conda/lib/python3.7/site-packages (from dataprep) (1.8.1)
Requirement already satisfied: pandas<2.0,>=1.1 in /opt/conda/lib/python3.7/site-packages (from dataprep) (1.2.3)
Collecting jsonpath-ng<2.0,>=1.5
  Downloading jsonpath_ng-1.5.3-py3-none-any.whl (29 kB)
Collecting dask[array,dataframe,delayed]<3.0,>=2.25
  Downloading dask-2.30.0-py3-none-any.whl (848 kB)
     |████████████████████████████████| 848 kB 4.7 MB/s eta 0:00:01
Collecting bottleneck<2.0,>=1.3
  Downloading Bottleneck-1.3.2.tar.gz (88 kB)
     |████████████████████████████████| 88 kB 2.8 MB/s eta 0:00:01
  Installing build dependencies ... done
  Getting requirements to build wheel ... done
    Preparing wheel metadata ... done
Collecting regex<2021.0.0,>=2020.10.15
  Downloading regex-2020.11.13-cp37-cp37m-manylinux2014_x86_64.whl (719 kB)
     |████████████████████████████████| 719 kB 5.3 MB/s eta 0:00:01
Collecting nltk<4.0,>=3.5
  Downloading nltk-3.6.2-py3-none-any.whl (1.5 MB)
     |████████████████████████████████| 1.5 MB 5.2 MB/s eta 0:00:01
Requirement already satisfied: bokeh<3,>=2 in /opt/conda/lib/python3.7/site-packages (from dataprep) (2.3.1)
Collecting pydantic<2.0,>=1.6
  Downloading pydantic-1.8.2-cp37-cp37m-manylinux2014_x86_64.whl (10.1 MB)
     |████████████████████████████████| 10.1 MB 3.1 MB/s eta 0:00:01
Requirement already satisfied: scipy<2,>=1 in /opt/conda/lib/python3.7/site-packages (from dataprep) (1.5.4)
Requirement already satisfied: python-Levenshtein<0.13.0,>=0.12.2 in /opt/conda/lib/python3.7/site-packages (from dataprep) (0.12.
2)
Collecting metaphone<0.7,>=0.6
  Downloading Metaphone-0.6.tar.gz (14 kB)
Requirement already satisfied: aiohttp<4.0,>=3.6 in /opt/conda/lib/python3.7/site-packages (from dataprep) (3.7.4)
Requirement already satisfied: typing-extensions>=3.6.5 in /opt/conda/lib/python3.7/site-packages (from aiohttp<4.0,>=3.6->datapre
p) (3.7.4.3)
Requirement already satisfied: chardet<5.0,>=2.0 in /opt/conda/lib/python3.7/site-packages (from aiohttp<4.0,>=3.6->dataprep) (4.
0.0)
Requirement already satisfied: multidict<7.0,>=4.5 in /opt/conda/lib/python3.7/site-packages (from aiohttp<4.0,>=3.6->dataprep)
(5.1.0)
```

```
Requirement already satisfied: yarl<2.0,>=1.0 in /opt/conda/lib/python3.7/site-packages (from aiohttp<4.0,>=3.6->dataprep) (1.6.3)
Requirement already satisfied: attrs>=17.3.0 in /opt/conda/lib/python3.7/site-packages (from aiohttp<4.0,>=3.6->dataprep) (20.3.0)
Requirement already satisfied: async-timeout<4.0,>=3.0 in /opt/conda/lib/python3.7/site-packages (from aiohttp<4.0,>=3.6->datapre
p) (3.0.1)
Requirement already satisfied: python-dateutil>=2.1 in /opt/conda/lib/python3.7/site-packages (from bokeh<3,>=2->dataprep) (2.8.1)
Requirement already satisfied: packaging>=16.8 in /opt/conda/lib/python3.7/site-packages (from bokeh<3,>=2->dataprep) (20.9)
Requirement already satisfied: pillow>=7.1.0 in /opt/conda/lib/python3.7/site-packages (from bokeh<3,>=2->dataprep) (7.2.0)
Requirement already satisfied: PyYAML>=3.10 in /opt/conda/lib/python3.7/site-packages (from bokeh<3,>=2->dataprep) (5.3.1)
Requirement already satisfied: tornado>=5.1 in /opt/conda/lib/python3.7/site-packages (from bokeh<3,>=2->dataprep) (6.1)
Requirement already satisfied: toolz>=0.8.2 in /opt/conda/lib/python3.7/site-packages (from dask[array,dataframe,delayed]<3.0,>=2.
25->dataprep) (0.11.1)
Requirement already satisfied: cloudpickle>=0.2.2 in /opt/conda/lib/python3.7/site-packages (from dask[array,dataframe,delayed]<3.
0,>=2.25->dataprep) (1.6.0)
Requirement already satisfied: partd>=0.3.10 in /opt/conda/lib/python3.7/site-packages (from dask[array,dataframe,delayed]<3.0,>=
2.25->dataprep) (1.2.0)
Requirement already satisfied: fsspec>=0.6.0 in /opt/conda/lib/python3.7/site-packages (from dask[array,dataframe,delayed]<3.0,>=
2.25->dataprep) (0.8.7)
Requirement already satisfied: importlib-metadata in /opt/conda/lib/python3.7/site-packages (from fsspec>=0.6.0->dask[array,datafr
ame,delayed]<3.0,>=2.25->dataprep) (3.4.0)
Requirement already satisfied: nbformat>=4.2.0 in /opt/conda/lib/python3.7/site-packages (from ipywidgets<8.0,>=7.5->dataprep) (5.
1.2)
Requirement already satisfied: ipython>=4.0.0 in /opt/conda/lib/python3.7/site-packages (from ipywidgets<8.0,>=7.5->dataprep) (7.2
2.0)
Requirement already satisfied: widgetsnbextension~=3.5.0 in /opt/conda/lib/python3.7/site-packages (from ipywidgets<8.0,>=7.5->dat
aprep) (3.5.1)
Requirement already satisfied: ipykernel>=4.5.1 in /opt/conda/lib/python3.7/site-packages (from ipywidgets<8.0,>=7.5->dataprep)
(5.5.0)
Requirement already satisfied: jupyterlab-widgets>=1.0.0 in /opt/conda/lib/python3.7/site-packages (from ipywidgets<8.0,>=7.5->dat
aprep) (1.0.0)
Requirement already satisfied: traitlets>=4.3.1 in /opt/conda/lib/python3.7/site-packages (from ipywidgets<8.0,>=7.5->dataprep)
(5.0.5)
Requirement already satisfied: jupyter-client in /opt/conda/lib/python3.7/site-packages (from ipykernel>=4.5.1->ipywidgets<8.0,>=
7.5->dataprep) (6.1.12)
Requirement already satisfied: prompt-toolkit!=3.0.0,!=3.0.1,<3.1.0,>=2.0.0 in /opt/conda/lib/python3.7/site-packages (from ipytho
n>=4.0.0->ipywidgets<8.0,>=7.5->dataprep) (3.0.18)
Requirement already satisfied: setuptools>=18.5 in /opt/conda/lib/python3.7/site-packages (from ipython>=4.0.0->ipywidgets<8.0,>=
7.5->dataprep) (49.6.0.post20210108)
Requirement already satisfied: pickleshare in /opt/conda/lib/python3.7/site-packages (from ipython>=4.0.0->ipywidgets<8.0,>=7.5->d
ataprep) (0.7.5)
Requirement already satisfied: pygments in /opt/conda/lib/python3.7/site-packages (from ipython>=4.0.0->ipywidgets<8.0,>=7.5->data
prep) (2.8.1)
Requirement already satisfied: decorator in /opt/conda/lib/python3.7/site-packages (from ipython>=4.0.0->ipywidgets<8.0,>=7.5->dat
aprep) (4.4.2)
Requirement already satisfied: backcall in /opt/conda/lib/python3.7/site-packages (from ipython>=4.0.0->ipywidgets<8.0,>=7.5->data
prep) (0.2.0)
Requirement already satisfied: pexpect>4.3 in /opt/conda/lib/python3.7/site-packages (from ipython>=4.0.0->ipywidgets<8.0,>=7.5->d
ataprep) (4.8.0)
```

```
Requirement already satisfied: jedi>=0.16 in /opt/conda/lib/python3.7/site-packages (from ipython>=4.0.0->ipywidgets<8.0,>=7.5->da
taprep) (0.18.0)
Requirement already satisfied: parso<0.9.0,>=0.8.0 in /opt/conda/lib/python3.7/site-packages (from jedi>=0.16->ipython>=4.0.0->ipy
widgets<8.0,>=7.5->dataprep) (0.8.1)
Requirement already satisfied: MarkupSafe>=0.23 in /opt/conda/lib/python3.7/site-packages (from jinja2<3.0,>=2.11->dataprep) (1.1.
1)
Requirement already satisfied: six in /opt/conda/lib/python3.7/site-packages (from jsonpath-ng<2.0,>=1.5->dataprep) (1.15.0)
Collecting ply
  Downloading ply-3.11-py2.py3-none-any.whl (49 kB)
     |████████████████████████████████| 49 kB 2.5 MB/s eta 0:00:011
Requirement already satisfied: jupyter-core in /opt/conda/lib/python3.7/site-packages (from nbformat>=4.2.0->ipywidgets<8.0,>=7.5-
>dataprep) (4.7.1)
Requirement already satisfied: jsonschema!=2.5.0,>=2.4 in /opt/conda/lib/python3.7/site-packages (from nbformat>=4.2.0->ipywidgets
<8.0,>=7.5->dataprep) (3.2.0)
Requirement already satisfied: ipython-genutils in /opt/conda/lib/python3.7/site-packages (from nbformat>=4.2.0->ipywidgets<8.0,>=
7.5->dataprep) (0.2.0)
Requirement already satisfied: pyrsistent>=0.14.0 in /opt/conda/lib/python3.7/site-packages (from jsonschema!=2.5.0,>=2.4->nbforma
t>=4.2.0->ipywidgets<8.0,>=7.5->dataprep) (0.17.3)
Requirement already satisfied: click in /opt/conda/lib/python3.7/site-packages (from nltk<4.0,>=3.5->dataprep) (7.1.2)
Requirement already satisfied: joblib in /opt/conda/lib/python3.7/site-packages (from nltk<4.0,>=3.5->dataprep) (1.0.1)
Requirement already satisfied: pyparsing>=2.0.2 in /opt/conda/lib/python3.7/site-packages (from packaging>=16.8->bokeh<3,>=2->data
prep) (2.4.7)
Requirement already satisfied: pytz>=2017.3 in /opt/conda/lib/python3.7/site-packages (from pandas<2.0,>=1.1->dataprep) (2021.1)
Requirement already satisfied: locket in /opt/conda/lib/python3.7/site-packages (from partd>=0.3.10->dask[array,dataframe,delayed]
<3.0,>=2.25->dataprep) (0.2.1)
Requirement already satisfied: ptyprocess>=0.5 in /opt/conda/lib/python3.7/site-packages (from pexpect>4.3->ipython>=4.0.0->ipywid
gets<8.0,>=7.5->dataprep) (0.7.0)
Requirement already satisfied: wcwidth in /opt/conda/lib/python3.7/site-packages (from prompt-toolkit!=3.0.0,!=3.0.1,<3.1.0,>=2.0.
0->ipython>=4.0.0->ipywidgets<8.0,>=7.5->dataprep) (0.2.5)
Requirement already satisfied: future>=0.14 in /opt/conda/lib/python3.7/site-packages (from usaddress<0.6.0,>=0.5.10->dataprep)
(0.18.2)
Collecting probableparsing
  Downloading probableparsing-0.0.1-py2.py3-none-any.whl (3.1 kB)
Collecting python-crfsuite>=0.7
  Downloading python_crfsuite-0.9.7-cp37-cp37m-manylinux1_x86_64.whl (743 kB)
     |████████████████████████████████| 743 kB 5.5 MB/s eta 0:00:01
Requirement already satisfied: notebook>=4.4.1 in /opt/conda/lib/python3.7/site-packages (from widgetsnbextension~=3.5.0->ipywidge
ts<8.0,>=7.5->dataprep) (6.3.0)
Requirement already satisfied: nbconvert in /opt/conda/lib/python3.7/site-packages (from notebook>=4.4.1->widgetsnbextension~=3.5.
0->ipywidgets<8.0,>=7.5->dataprep) (6.0.7)
Requirement already satisfied: argon2-cffi in /opt/conda/lib/python3.7/site-packages (from notebook>=4.4.1->widgetsnbextension~=3.
5.0->ipywidgets<8.0,>=7.5->dataprep) (20.1.0)
Requirement already satisfied: pyzmq>=17 in /opt/conda/lib/python3.7/site-packages (from notebook>=4.4.1->widgetsnbextension~=3.5.
0->ipywidgets<8.0,>=7.5->dataprep) (22.0.3)
Requirement already satisfied: terminado>=0.8.3 in /opt/conda/lib/python3.7/site-packages (from notebook>=4.4.1->widgetsnbextensio
n~=3.5.0->ipywidgets<8.0,>=7.5->dataprep) (0.9.3)
Requirement already satisfied: prometheus-client in /opt/conda/lib/python3.7/site-packages (from notebook>=4.4.1->widgetsnbextensi
```
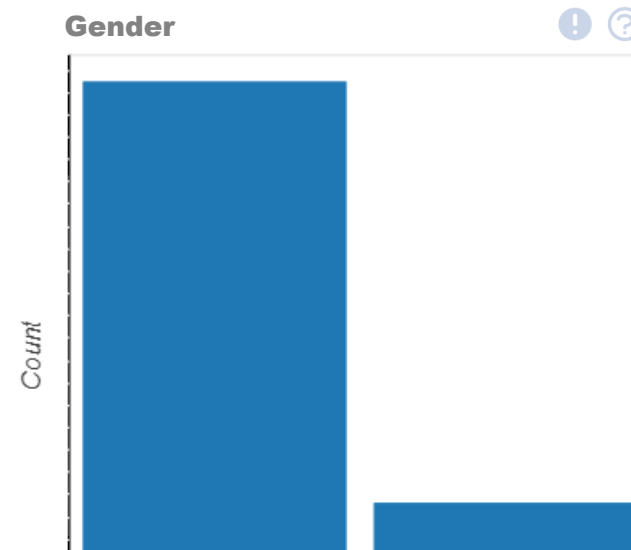
```
on~=3.5.0->ipywidgets<8.0,>=7.5->dataprep) (0.9.0)
Requirement already satisfied: Send2Trash>=1.5.0 in /opt/conda/lib/python3.7/site-packages (from notebook>=4.4.1->widgetsnbextensi
on~=3.5.0->ipywidgets<8.0,>=7.5->dataprep) (1.5.0)
Requirement already satisfied: matplotlib in /opt/conda/lib/python3.7/site-packages (from wordcloud<2.0,>=1.8->dataprep) (3.4.1)
Requirement already satisfied: idna>=2.0 in /opt/conda/lib/python3.7/site-packages (from yarl<2.0,>=1.0->aiohttp<4.0,>=3.6->datapr
ep) (2.10)
Requirement already satisfied: cffi>=1.0.0 in /opt/conda/lib/python3.7/site-packages (from argon2-cffi->notebook>=4.4.1->widgetsnb
extension~=3.5.0->ipywidgets<8.0,>=7.5->dataprep) (1.14.5)
Requirement already satisfied: pycparser in /opt/conda/lib/python3.7/site-packages (from cffi>=1.0.0->argon2-cffi->notebook>=4.4.1
->widgetsnbextension~=3.5.0->ipywidgets<8.0,>=7.5->dataprep) (2.20)
Requirement already satisfied: zipp>=0.5 in /opt/conda/lib/python3.7/site-packages (from importlib-metadata->fsspec>=0.6.0->dask[a
rray,dataframe,delayed]<3.0,>=2.25->dataprep) (3.4.1)
Requirement already satisfied: cycler>=0.10 in /opt/conda/lib/python3.7/site-packages (from matplotlib->wordcloud<2.0,>=1.8->datap
rep) (0.10.0)
Requirement already satisfied: kiwisolver>=1.0.1 in /opt/conda/lib/python3.7/site-packages (from matplotlib->wordcloud<2.0,>=1.8->
dataprep) (1.3.1)
Requirement already satisfied: defusedxml in /opt/conda/lib/python3.7/site-packages (from nbconvert->notebook>=4.4.1->widgetsnbext
ension~=3.5.0->ipywidgets<8.0,>=7.5->dataprep) (0.7.1)
Requirement already satisfied: bleach in /opt/conda/lib/python3.7/site-packages (from nbconvert->notebook>=4.4.1->widgetsnbextensi
on~=3.5.0->ipywidgets<8.0,>=7.5->dataprep) (3.3.0)
Requirement already satisfied: nbclient<0.6.0,>=0.5.0 in /opt/conda/lib/python3.7/site-packages (from nbconvert->notebook>=4.4.1->
widgetsnbextension~=3.5.0->ipywidgets<8.0,>=7.5->dataprep) (0.5.3)
Requirement already satisfied: pandocfilters>=1.4.1 in /opt/conda/lib/python3.7/site-packages (from nbconvert->notebook>=4.4.1->wi
dgetsnbextension~=3.5.0->ipywidgets<8.0,>=7.5->dataprep) (1.4.2)
Requirement already satisfied: jupyterlab-pygments in /opt/conda/lib/python3.7/site-packages (from nbconvert->notebook>=4.4.1->wid
getsnbextension~=3.5.0->ipywidgets<8.0,>=7.5->dataprep) (0.1.2)
Requirement already satisfied: entrypoints>=0.2.2 in /opt/conda/lib/python3.7/site-packages (from nbconvert->notebook>=4.4.1->widg
etsnbextension~=3.5.0->ipywidgets<8.0,>=7.5->dataprep) (0.3)
Requirement already satisfied: testpath in /opt/conda/lib/python3.7/site-packages (from nbconvert->notebook>=4.4.1->widgetsnbexten
sion~=3.5.0->ipywidgets<8.0,>=7.5->dataprep) (0.4.4)
Requirement already satisfied: mistune<2,>=0.8.1 in /opt/conda/lib/python3.7/site-packages (from nbconvert->notebook>=4.4.1->widge
tsnbextension~=3.5.0->ipywidgets<8.0,>=7.5->dataprep) (0.8.4)
Requirement already satisfied: async-generator in /opt/conda/lib/python3.7/site-packages (from nbclient<0.6.0,>=0.5.0->nbconvert->
notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets<8.0,>=7.5->dataprep) (1.10)
Requirement already satisfied: nest-asyncio in /opt/conda/lib/python3.7/site-packages (from nbclient<0.6.0,>=0.5.0->nbconvert->not
ebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets<8.0,>=7.5->dataprep) (1.4.3)
Requirement already satisfied: webencodings in /opt/conda/lib/python3.7/site-packages (from bleach->nbconvert->notebook>=4.4.1->wi
dgetsnbextension~=3.5.0->ipywidgets<8.0,>=7.5->dataprep) (0.5.1)
Building wheels for collected packages: bottleneck, metaphone
  Building wheel for bottleneck (PEP 517) ... done
  Created wheel for bottleneck: filename=Bottleneck-1.3.2-cp37-cp37m-linux_x86_64.whl size=334929 sha256=54504293fa66c85de2b14609b
2d1ca90f202e744426e9e6bece7e5a6632f39e9
  Stored in directory: /root/.cache/pip/wheels/87/85/9c/a325c89ff0498660ef8a335fb4b3912939c273ea4f094af29f
  Building wheel for metaphone (setup.py) ... done
  Created wheel for metaphone: filename=Metaphone-0.6-py3-none-any.whl size=13905 sha256=e76d899a52049da1280ecfda513eaf22724df952c
e9ee5fed22459acaea23bea
  Stored in directory: /root/.cache/pip/wheels/1d/a8/cb/6f8902aa5457bd71344e00665c230e9c45255b3f57f2194a0f
```
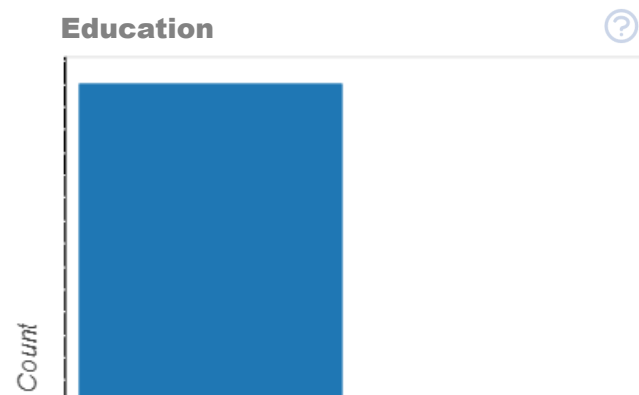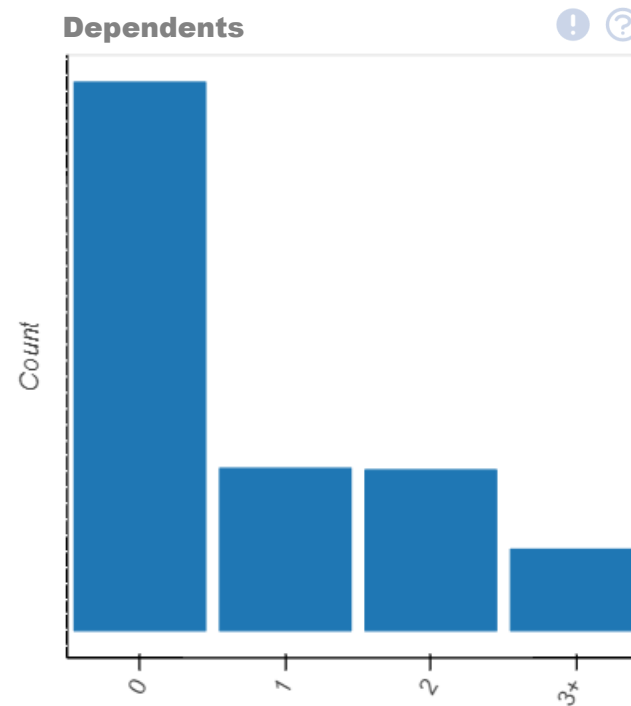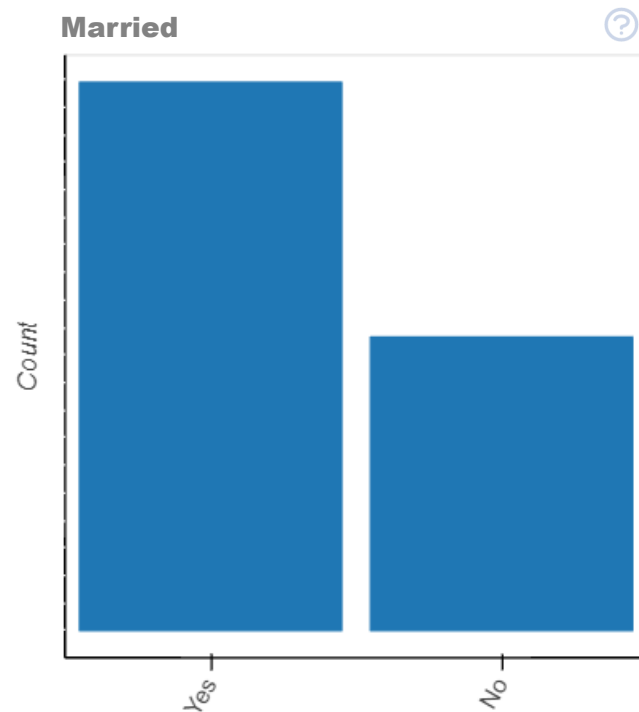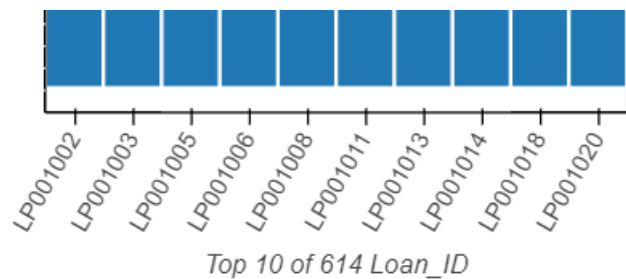
```
Successfully built bottleneck metaphone
Installing collected packages: regex, python-crfsuite, probableparsing, ply, dask, usaddress, pydantic, nltk, metaphone, jsonpath-
ng, bottleneck, dataprep
  Attempting uninstall: regex
    Found existing installation: regex 2021.3.17
    Uninstalling regex-2021.3.17:
      Successfully uninstalled regex-2021.3.17
  Attempting uninstall: dask
    Found existing installation: dask 2021.4.0
    Uninstalling dask-2021.4.0:
      Successfully uninstalled dask-2021.4.0
  Attempting uninstall: nltk
    Found existing installation: nltk 3.2.4
    Uninstalling nltk-3.2.4:
      Successfully uninstalled nltk-3.2.4
ERROR: pip's dependency resolver does not currently take into account all the packages that are installed. This behaviour is the s
ource of the following dependency conflicts.
preprocessing 0.1.13 requires nltk==3.2.4, but you have nltk 3.6.2 which is incompatible.
distributed 2021.4.0 requires dask>=2021.03.0, but you have dask 2.30.0 which is incompatible.
Successfully installed bottleneck-1.3.2 dask-2.30.0 dataprep-0.3.0 jsonpath-ng-1.5.3 metaphone-0.6 nltk-3.6.2 ply-3.11 probablepar
sing-0.0.1 pydantic-1.8.2 python-crfsuite-0.9.7 regex-2020.11.13 usaddress-0.5.10
```

In [6]:
```
plot(df_train)
```

Out[6]:

Show Stats and Insights

Top 10 of 614 Loan_ID

Married

Count

Yes   No

Dependents

Count

0   1   2   3+

Education

Count

Self_Employed

Count

Male   Female

ApplicantIncome

Frequency

0   20K   40K   60K   80K

CoapplicantIncome

Frequency

0   9000   18K   27K   36K

LoanAmount

Loan_Amount_Term

Graduate   Not Graduate

No   Yes
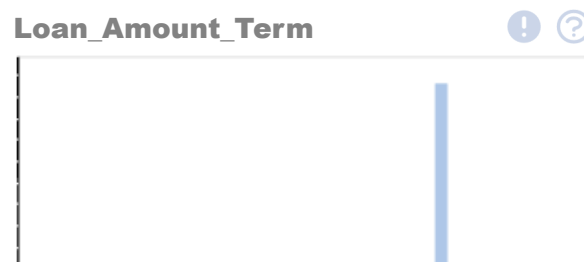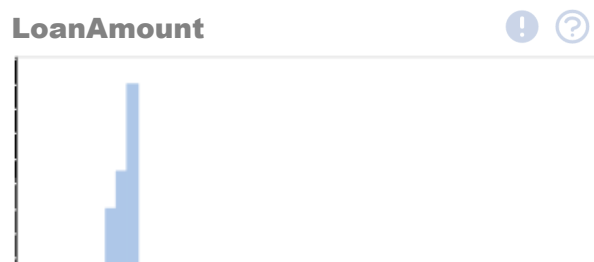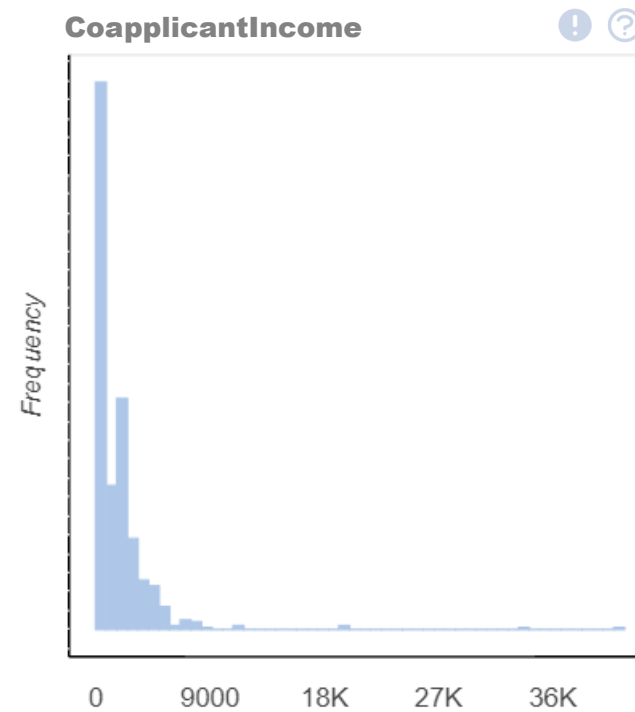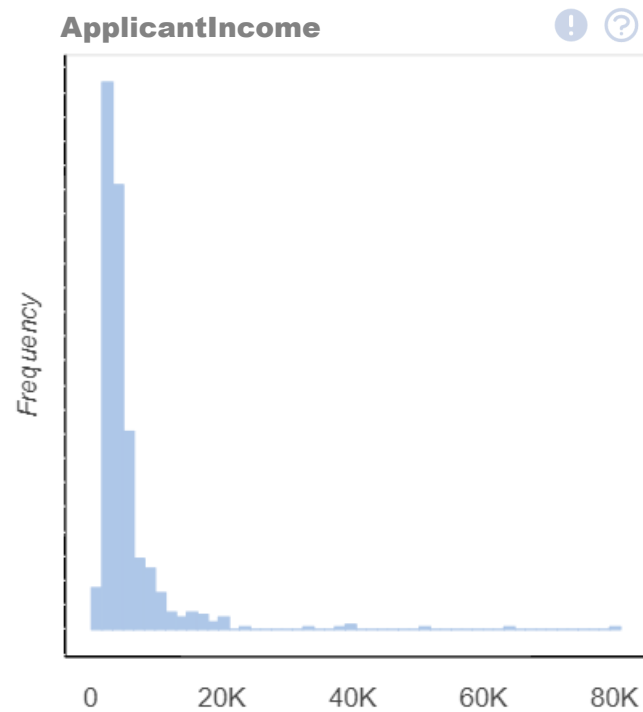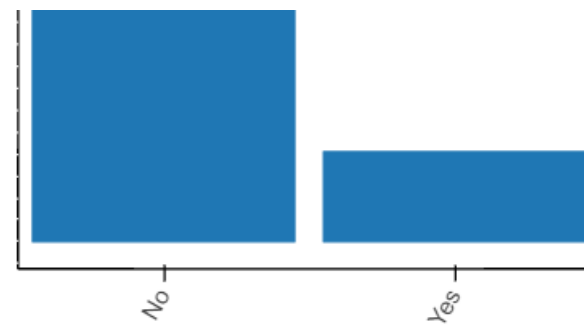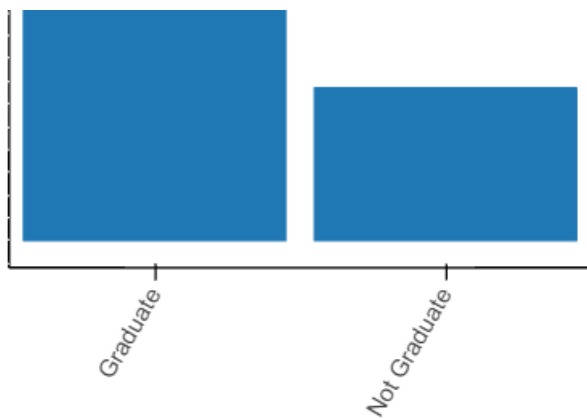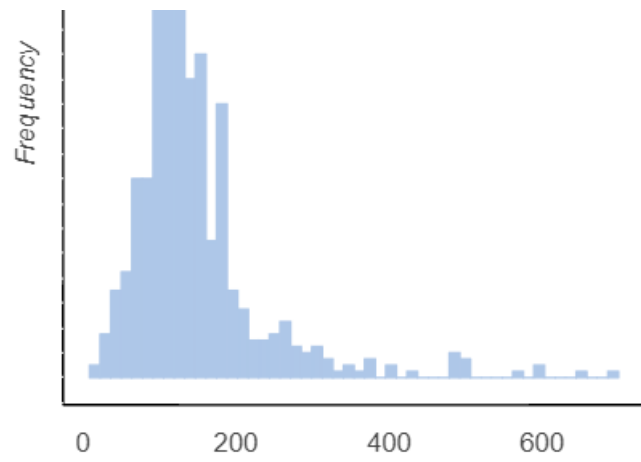
Credit_History

Property_Area

Loan_Status

Count

1   2
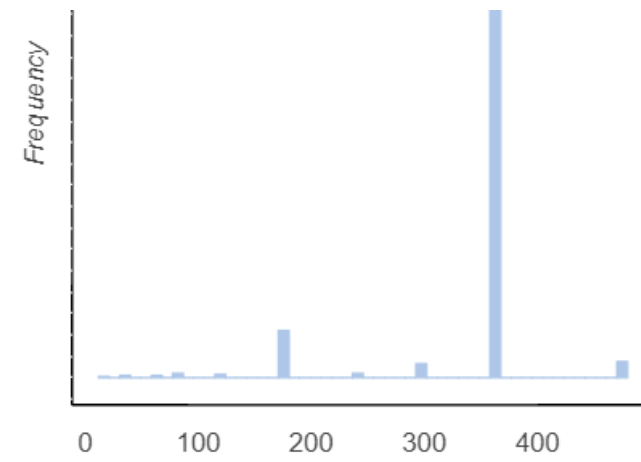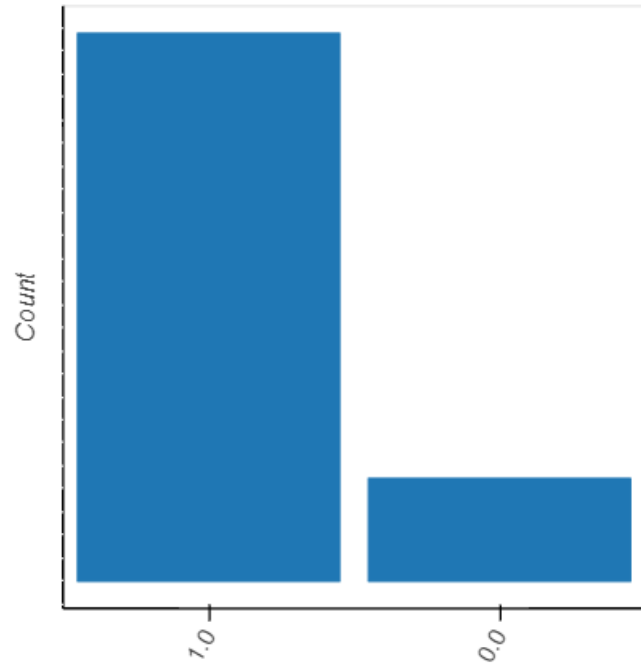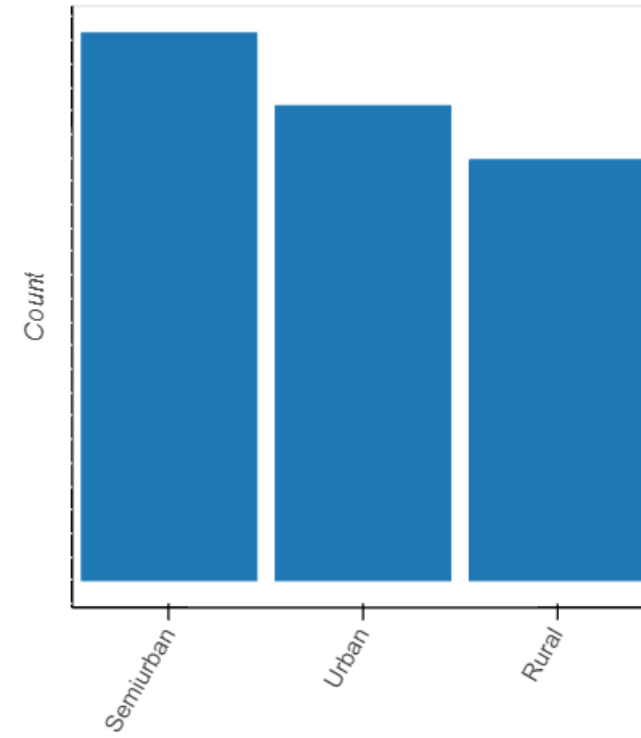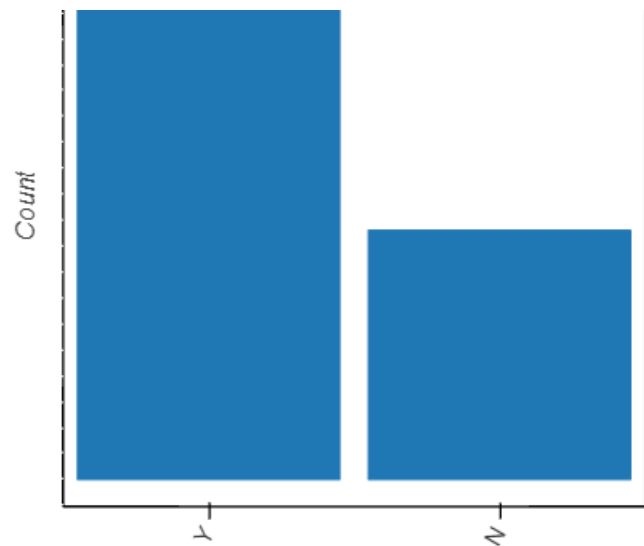
```
In [7]:  create_report(df_train)
```

Out[7]:

DataPrep Report   Overview   Variables ≡   Interactions   Correlations   Missing Values

Overview

Dataset Statistics

Number of Variables	13
Number of Rows	614
Missing Cells	149
Missing Cells (%)	1.9%
Duplicate Rows	0

Dataset Insights

`Gender` **has 13 (2.12%) missing values** `Missing`

`Dependents` **has 15 (2.44%) missing values** `Missing`

`Self_Employed` **has 32 (5.21%) missing values** `Missing`

`LoanAmount` **has 22 (3.58%) missing values** `Missing`

`Loan_Amount_Term` **has 14 (2.28%) missing values** `Missing`

Duplicate Rows (%)	0.0%
Total Size in Memory	316.6 KB
Average Row Size in Memory	528.0 B
Variable Types	Categorical: 8 GeoGraphy: 1 Numerical: 4

Credit_History has 50 (8.14%) missing values `Missing`

ApplicantIncome is skewed `Skewed`

CoapplicantIncome is skewed `Skewed`

LoanAmount is skewed `Skewed`

Loan_Amount_Term is skewed `Skewed`

`1` `2`

Variables

Loan_ID
categorical

Approximate Distinct Count	614
Approximate Unique (%)	**100.0%**
Missing	0
Missing (%)	0.0%
Memory Size	43.8 KB

`Show Details`

Loan_ID

Top 10 of 614 Loan_ID

Gender
categorical

Approximate Distinct Count	2
Approximate Unique (%)	0.3%
Missing	**13**

Gender

	Missing (%)	2.1%
	Memory Size	40.7 KB

Married
categorical

Approximate Distinct Count	2
Approximate Unique (%)	0.3%
Missing	3
Missing (%)	0.5%
Memory Size	40.4 KB
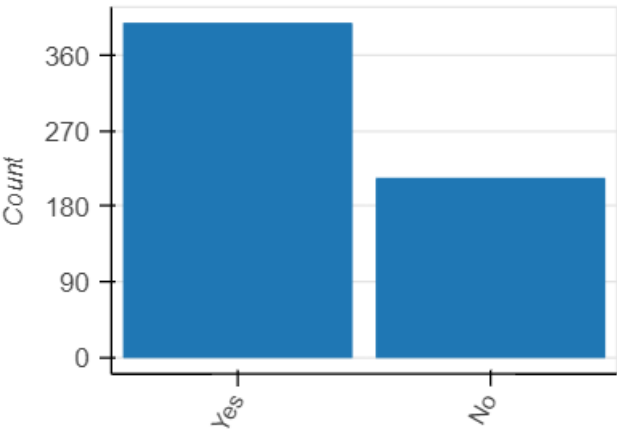
Show Details

Dependents
categorical

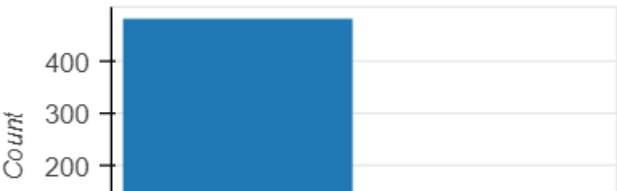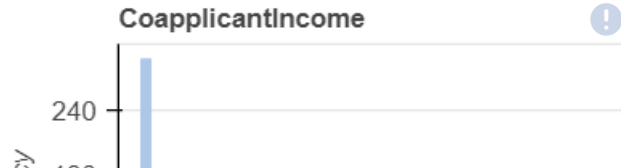| Approximate Distinct Count	4
Approximate Unique (%)	0.7%
Missing	15
Missing (%)	2.4%
Memory Size	38.7 KB
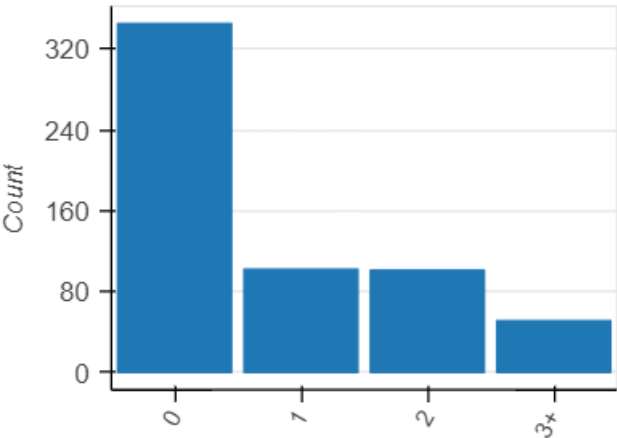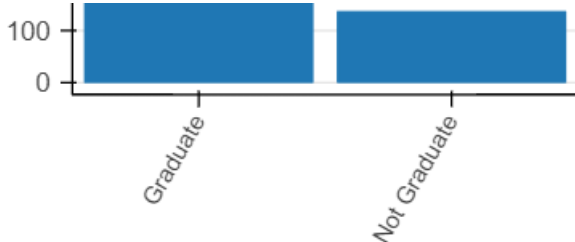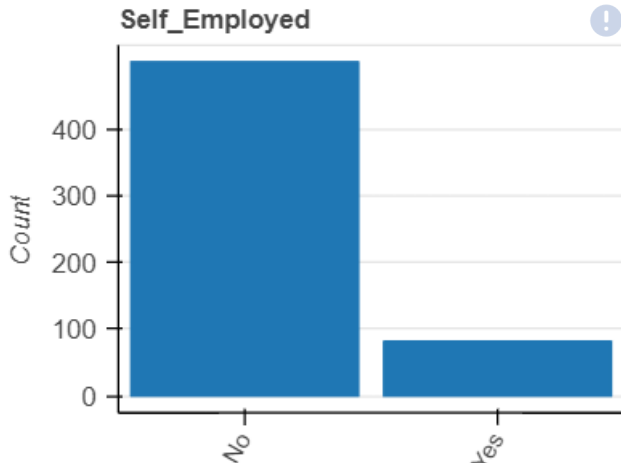
Show Details

Education

Approximate Distinct Count	2
Approximate Unique (%)	0.3%
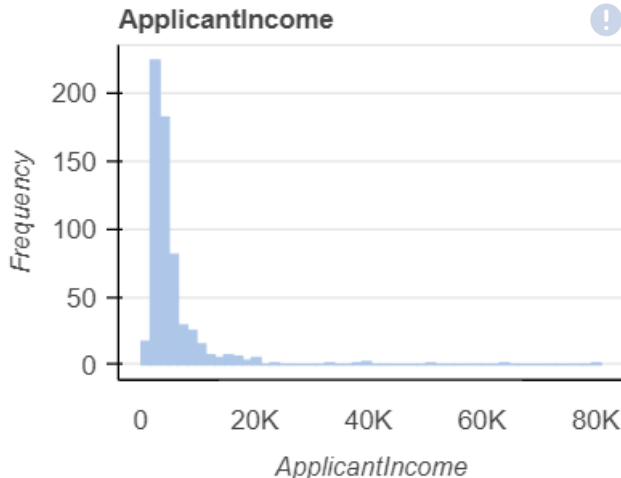
categorical	Missing	0
	Missing (%)	0.0%
Show Details	Memory Size	44.3 KB

Self_Employed

	Approximate Distinct Count	2
	Approximate Unique (%)	0.3%
Self_Employed	Missing	32
categorical	Missing (%)	5.2%
Show Details	Memory Size	38.2 KB

ApplicantIncome

	Approximate Distinct Count	505	Mean	5403.4593
	Approximate Unique (%)	82.2%	Minimum	150
ApplicantIncome			Maximum	81000
numerical	Missing	0	Zeros	0
	Missing (%)	0.0%	Zeros (%)	0.0%
Show Details	Infinite	0	Negatives	0
	Infinite (%)	0.0%	Negatives (%)	0.0%
	Memory Size	9.6 KB		

CoapplicantIncome

	Approximate Distinct Count	287	Missing	0
			Missing (%)	0.0%
	Approximate Unique (%)	46.7%	Infinite	0

CoapplicantIncome
numerical

[Show Details]

Infinite (%)	0.0%	Zeros	273
Memory Size	9.6 KB	Zeros (%)	44.5%
Mean	1621.2458	Negatives	0
Minimum	0	Negatives (%)	0.0%
Maximum	41667		
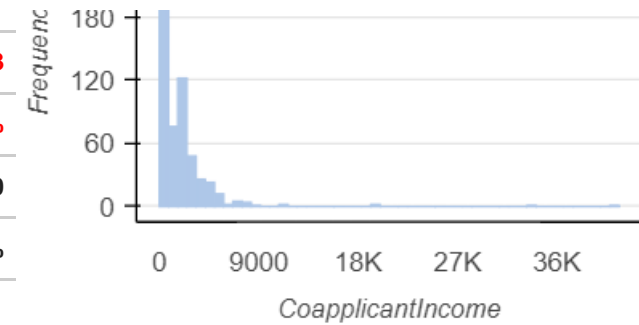
LoanAmount
numerical

[Show Details]

Approximate Distinct Count	203	Mean	146.4122
Approximate Unique (%)	34.3%	Minimum	9
		Maximum	700
Missing	22	Zeros	0
Missing (%)	3.6%	Zeros (%)	0.0%
Infinite	0	Negatives	0
Infinite (%)	0.0%	Negatives (%)	0.0%
Memory Size	9.2 KB		
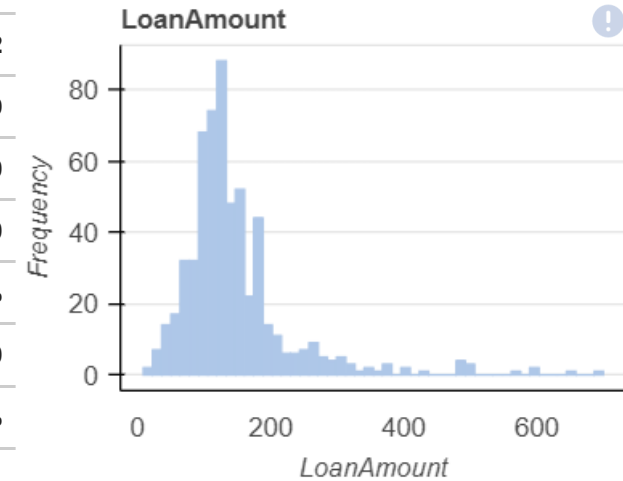
Loan_Amount_Term
numerical

[Show Details]

Approximate Distinct Count	10	Mean	342
Approximate Unique (%)	1.7%	Minimum	12
		Maximum	480
Missing	14	Zeros	0
Missing (%)	2.3%	Zeros (%)	0.0%
Infinite	0	Negatives	0
Infinite (%)	0.0%	Negatives (%)	0.0%
Memory Size	9.4 KB		
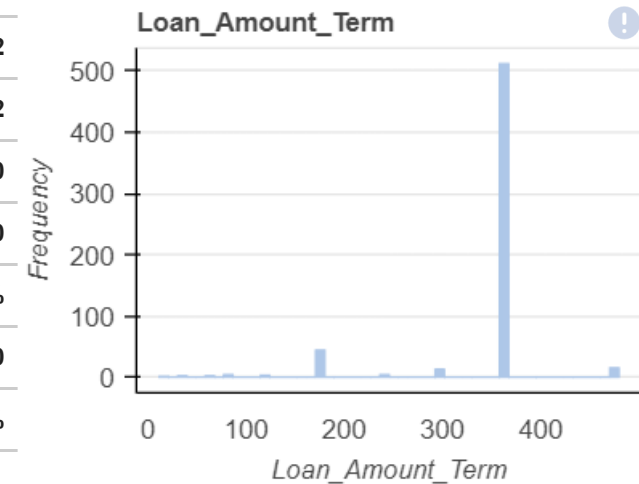
Credit_History

Approximate Distinct Count	2

Credit_History
categorical

[Show Details]

Approximate Unique (%)	0.4%
Missing	**50**
Missing (%)	**8.1%**
Memory Size	37.5 KB
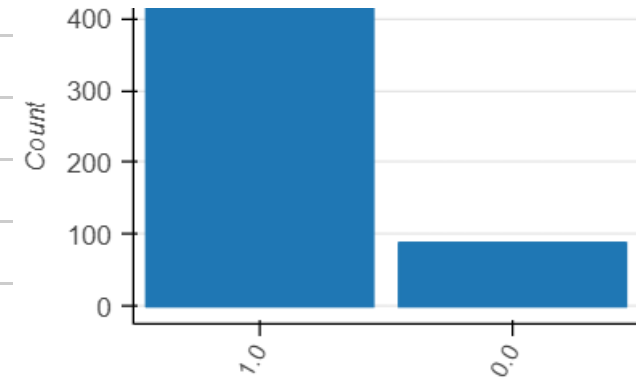
Property_Area
categorical

[Show Details]

Approximate Distinct Count	3
Approximate Unique (%)	0.5%
Missing	0
Missing (%)	0.0%
Memory Size	42.9 KB
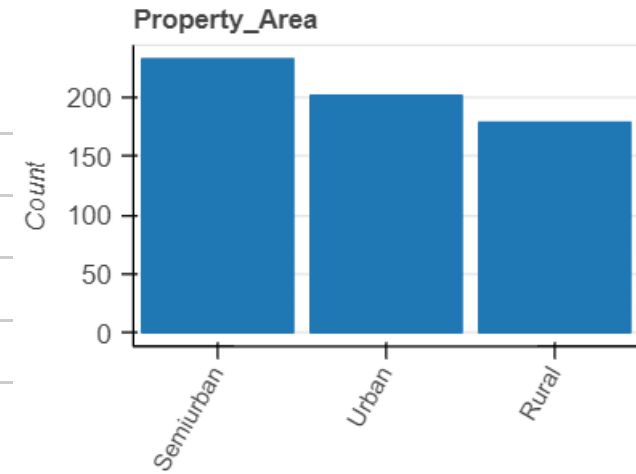
Property_Area

Loan_Status
categorical

[Show Details]

Approximate Distinct Count	2
Approximate Unique (%)	0.3%
Missing	0
Missing (%)	0.0%
Memory Size	39.6 KB

Loan_Status

Interactions

ApplicantIncome ▼   ApplicantIncome ▼

Scatter Plot

Correlations

Pearson	Spearman	KendallTau

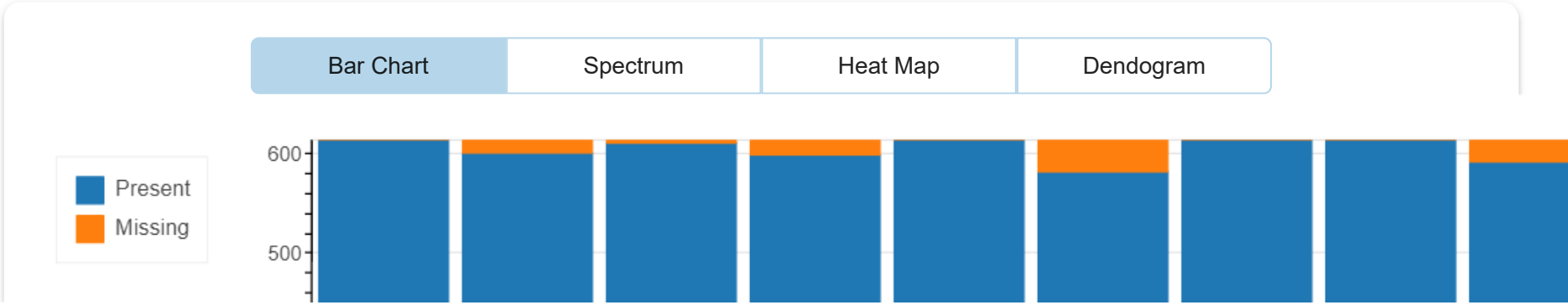## Missing Values

| Bar Chart | Spectrum | Heat Map | Dendogram |

Present
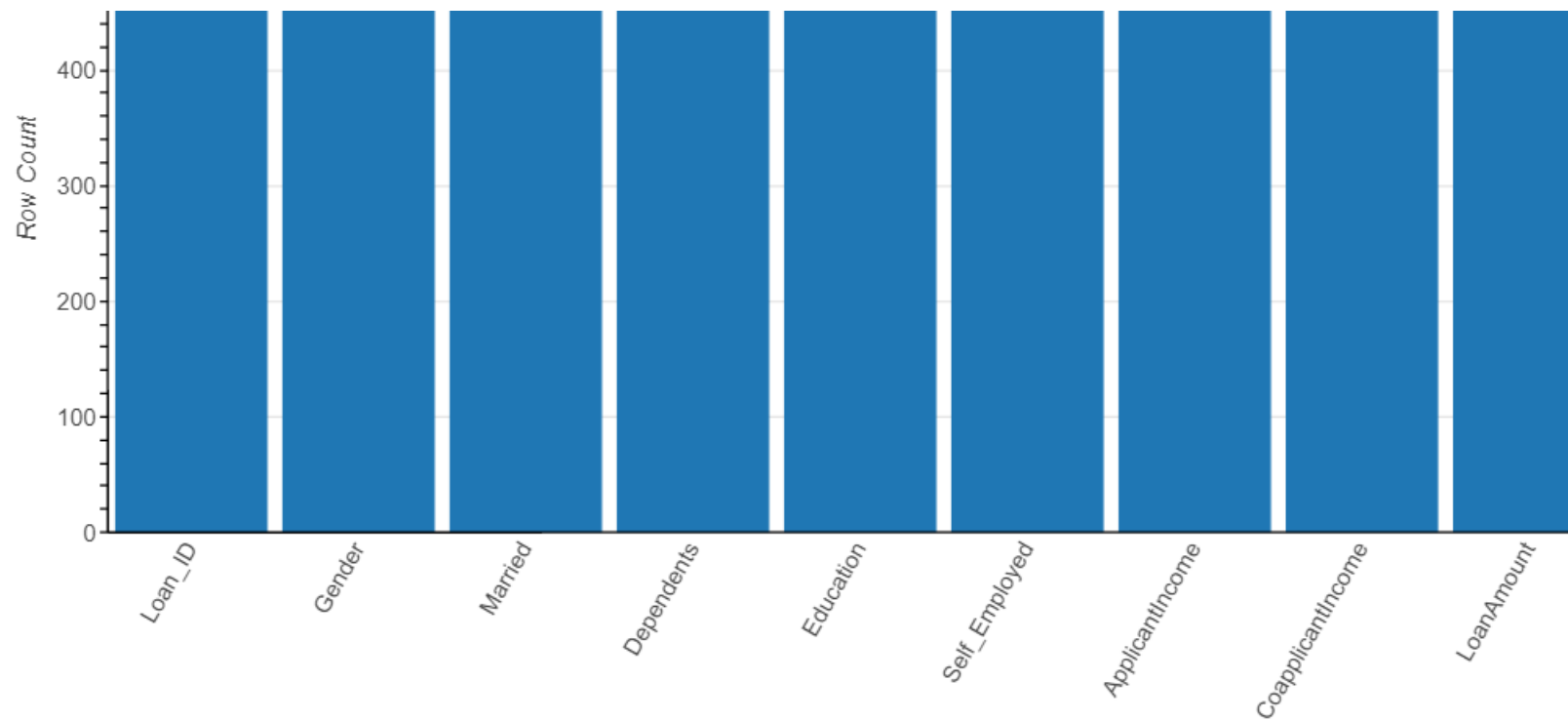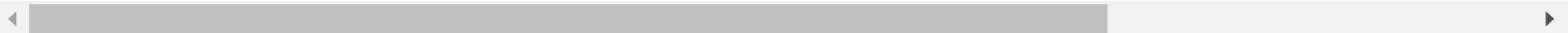Missing
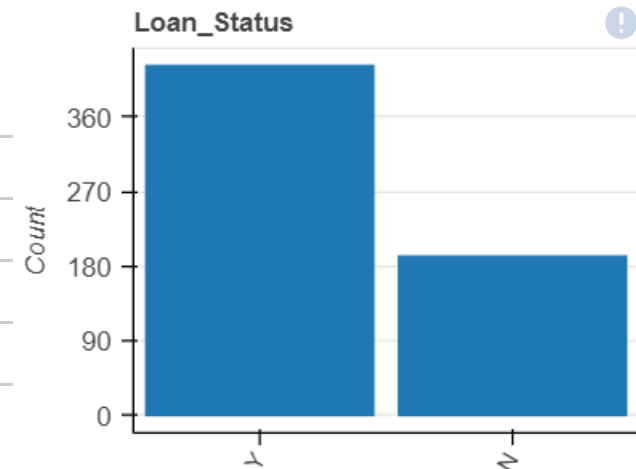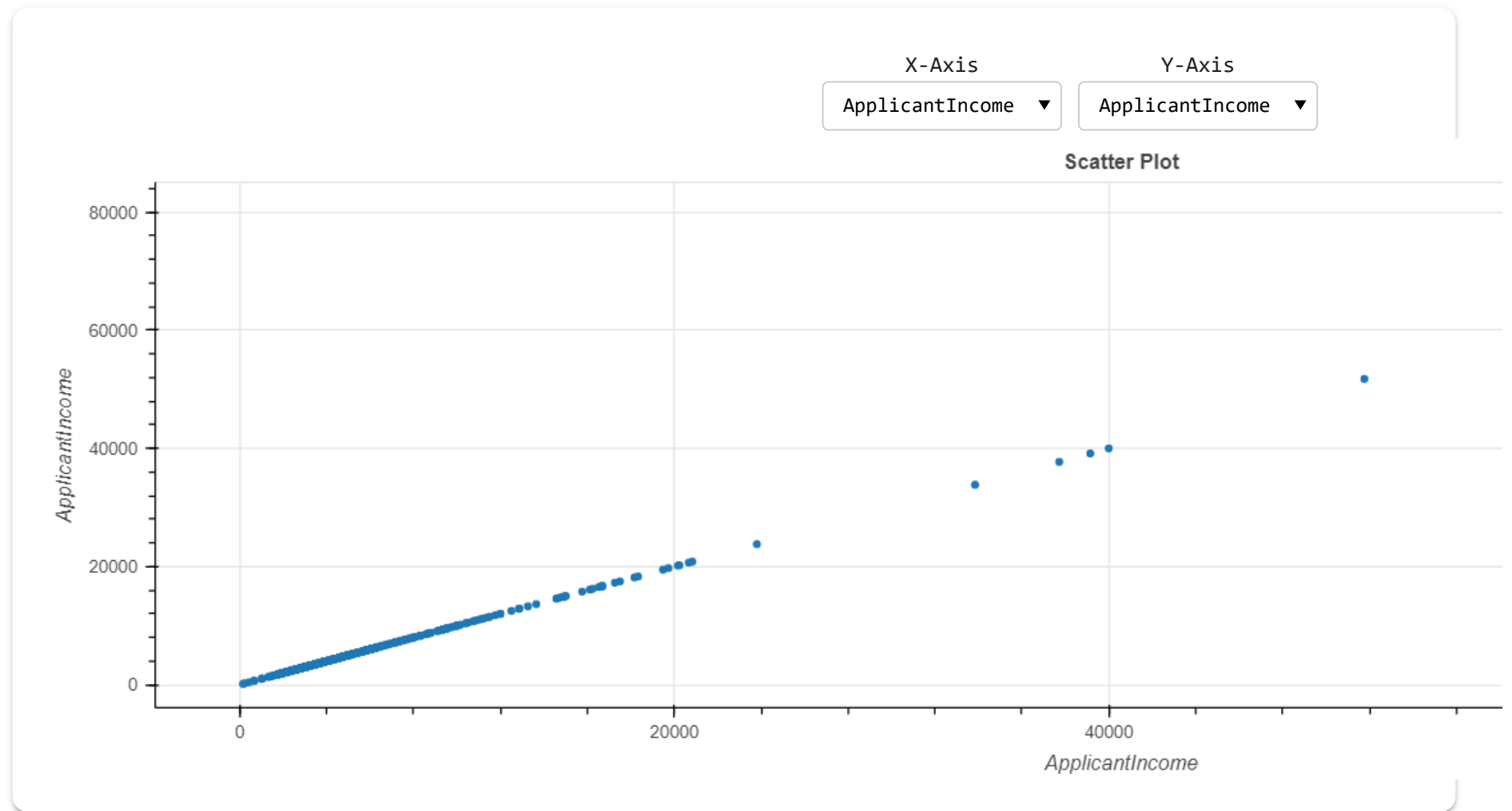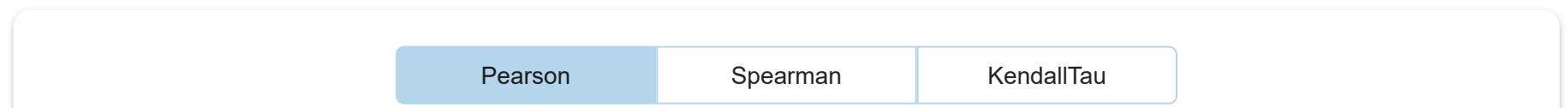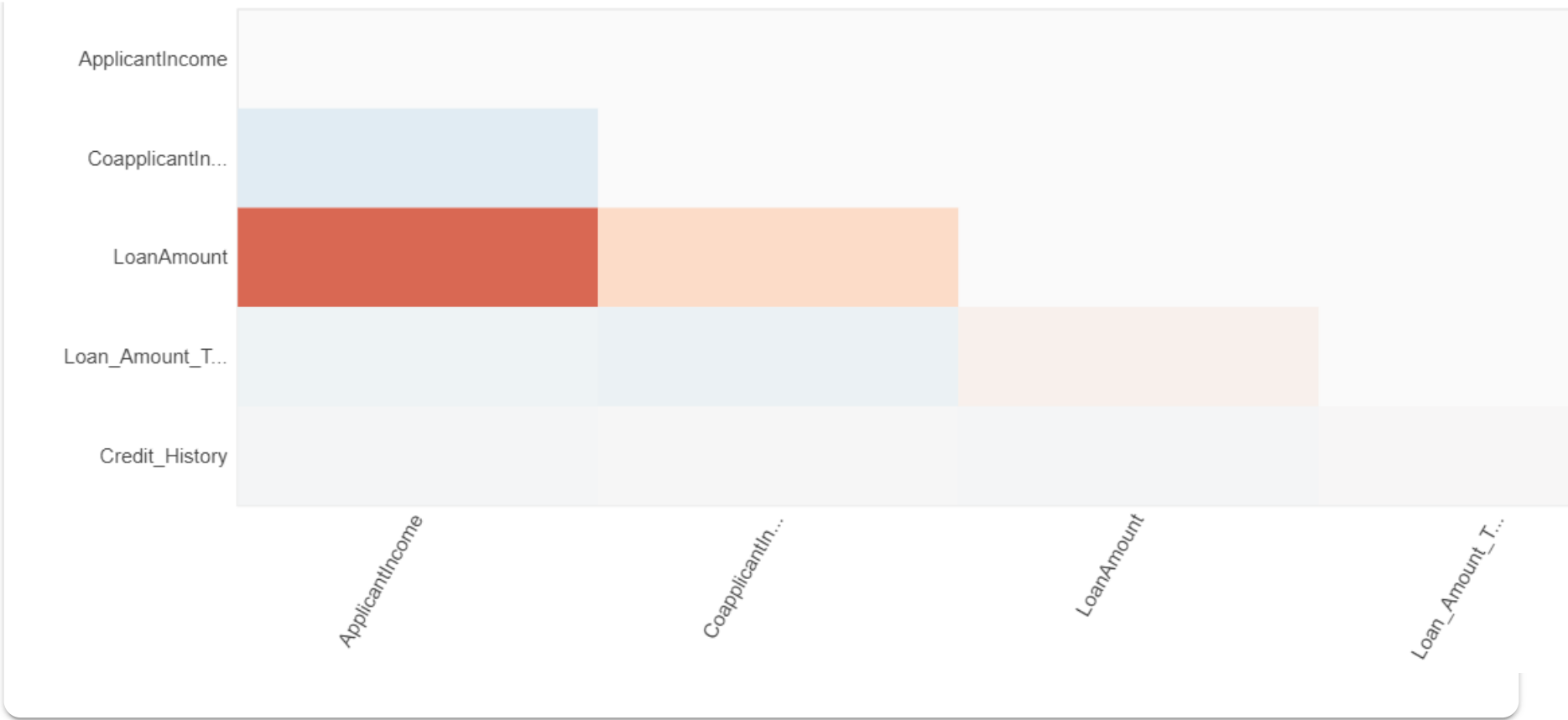
Report generated with DataPrep

In [8]: `plot(df_train, "Property_Area")`

Out[8]: Stats   Bar Chart   Pie Chart   Word Cloud   Word Frequency   Word Length   Value Table

Overview		Sample	
Approximate Distinct Count	3	**1st row**	Urban
Approximate Unique (%)	0.5%	**2nd row**	Rural

Missing	0	**3rd row**	Urban
Missing (%)	0.0%	**4th row**	Urban
Memory Size	42.9 KB	**5th row**	Urban

Length			Letter	
Mean	6.5179	**Count**	4002
Standard Deviation	1.9426	**Lowercase Letter**	3388
Median	5	**Space Separator**	0
Minimum	5	**Uppercase Letter**	614
Maximum	9	**Dash Punctuation**	0
		Decimal Number	0

In [ ]: