Neuro Data Engineering
Online EDA Automation

Wireframe documentation

Vamsi Alla

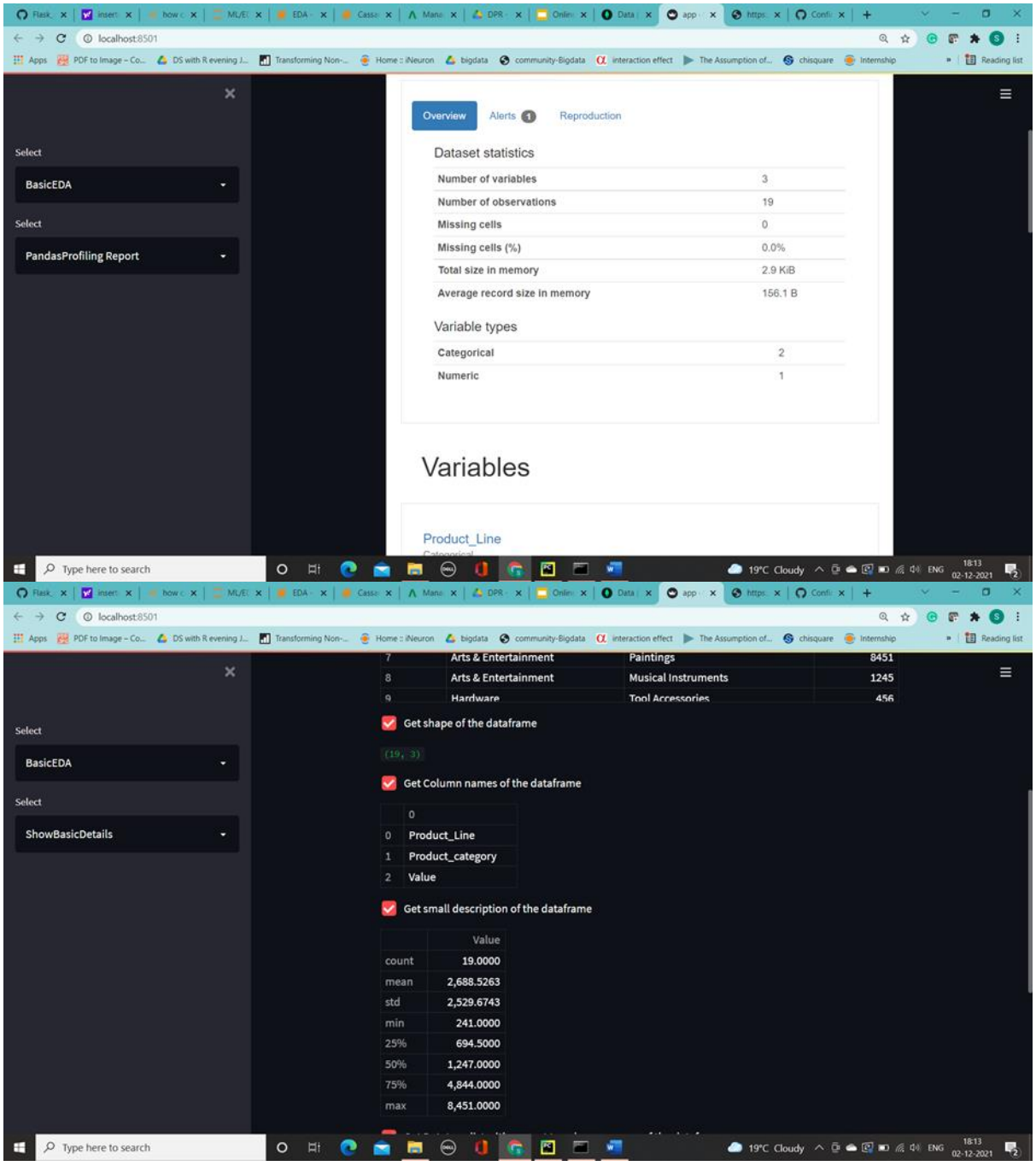
October 23, 2022

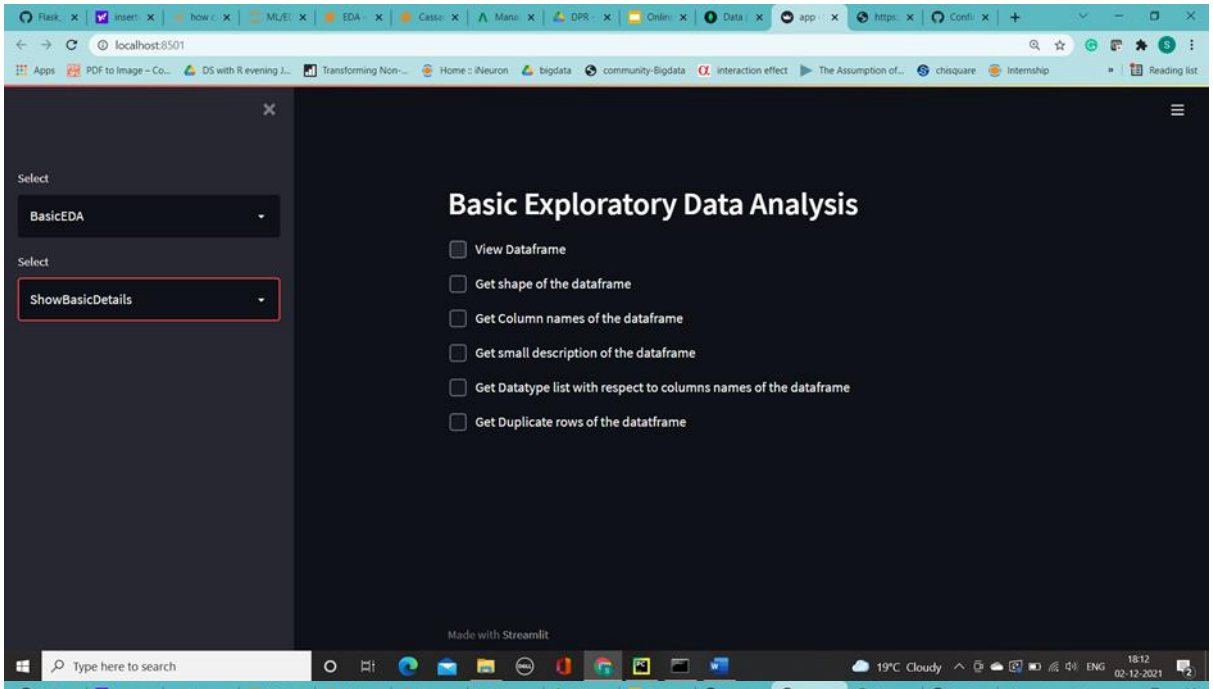
INEURON

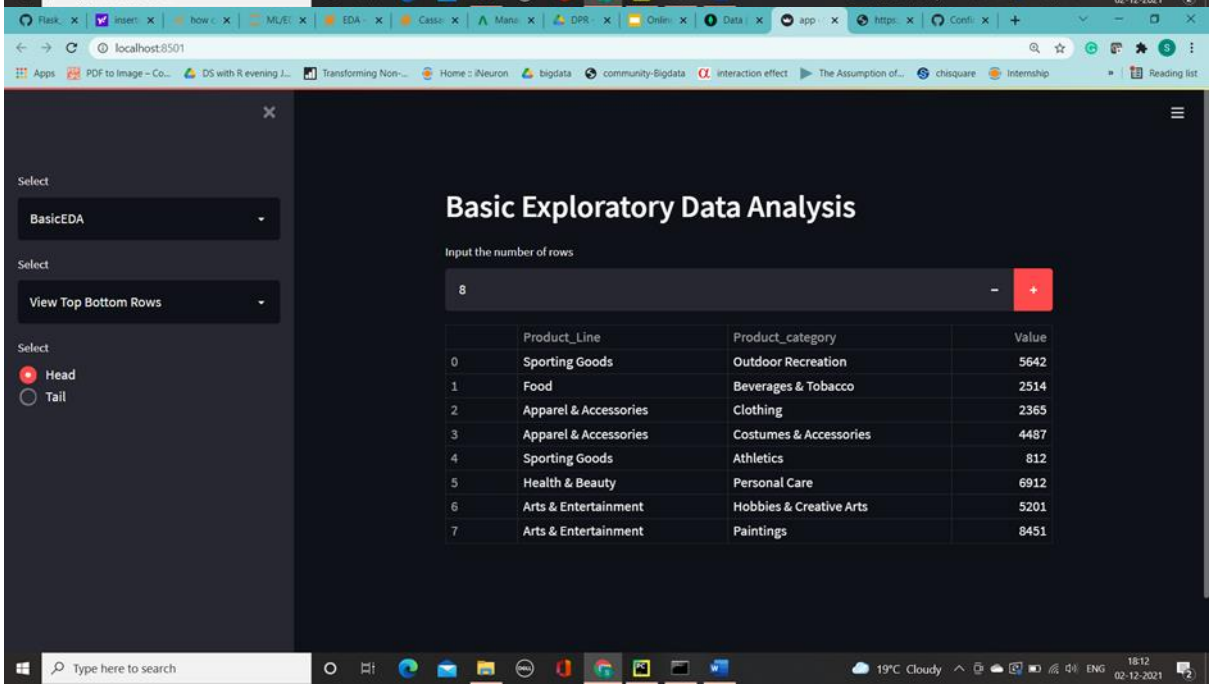
1. The first page displays the pop up window where the user has to provide specific details such as Basic EDA, Advanced EDA , Charts, Analysis of the dataset etc.

- All the information will be used to predict the results.

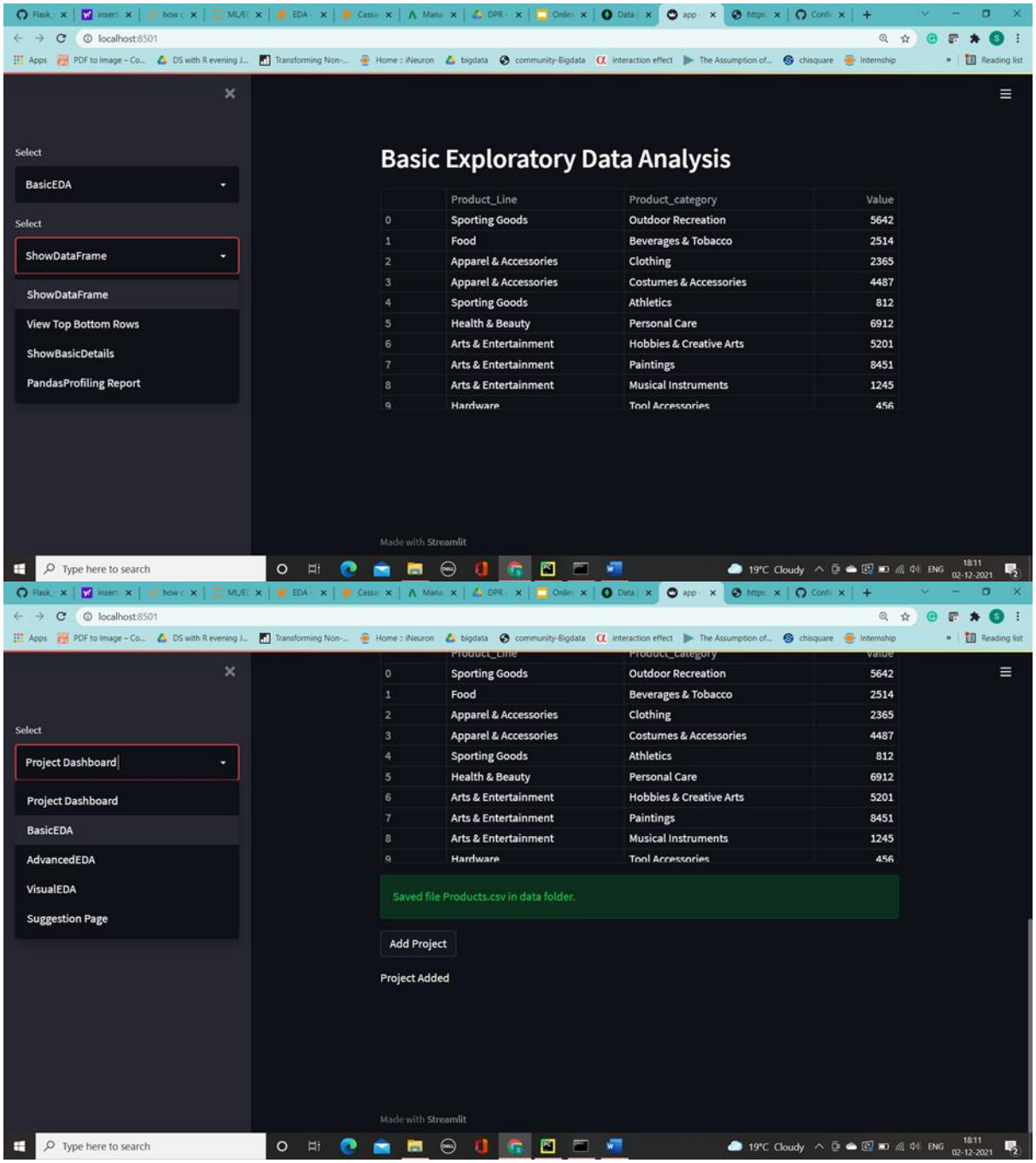
-

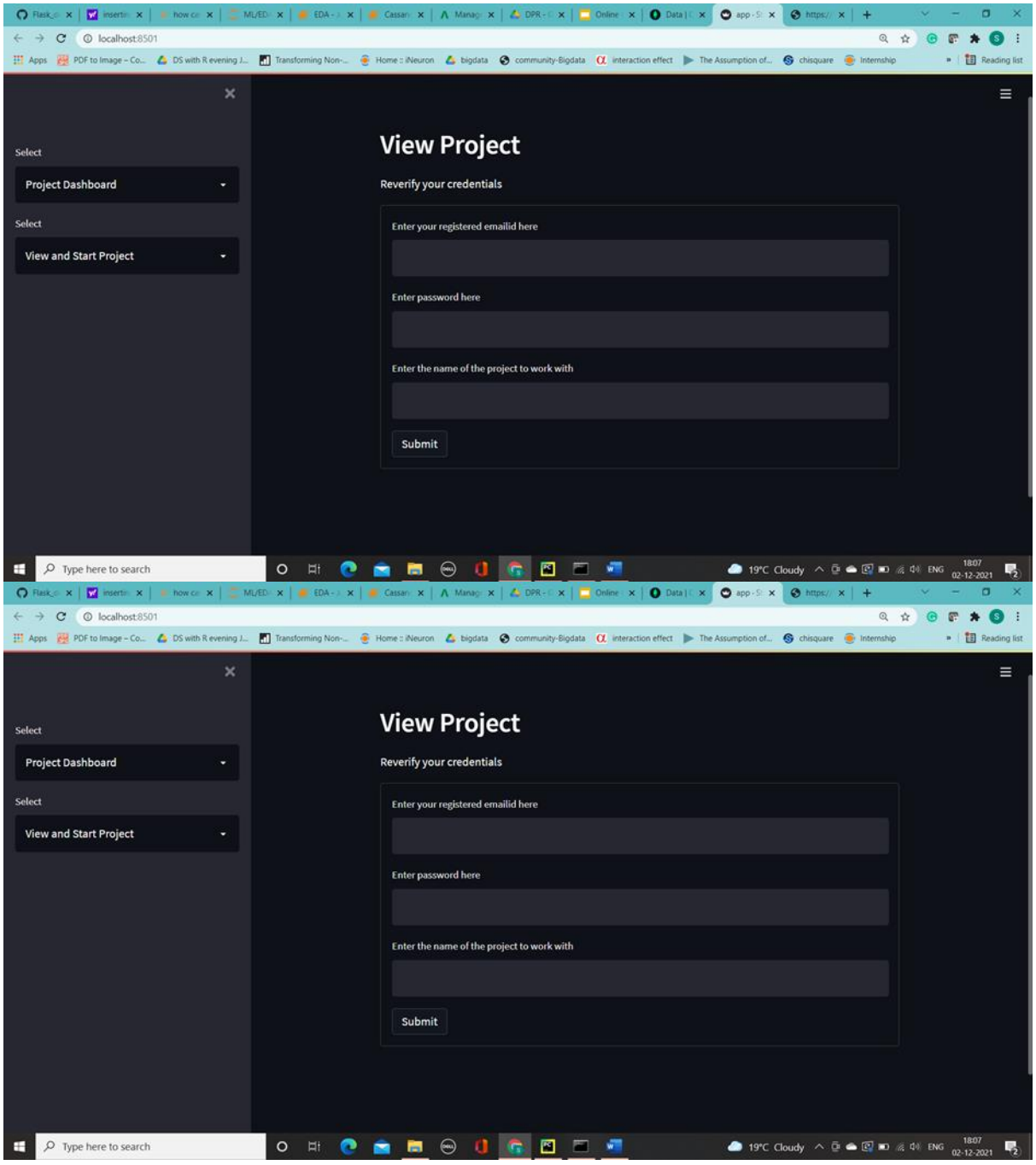




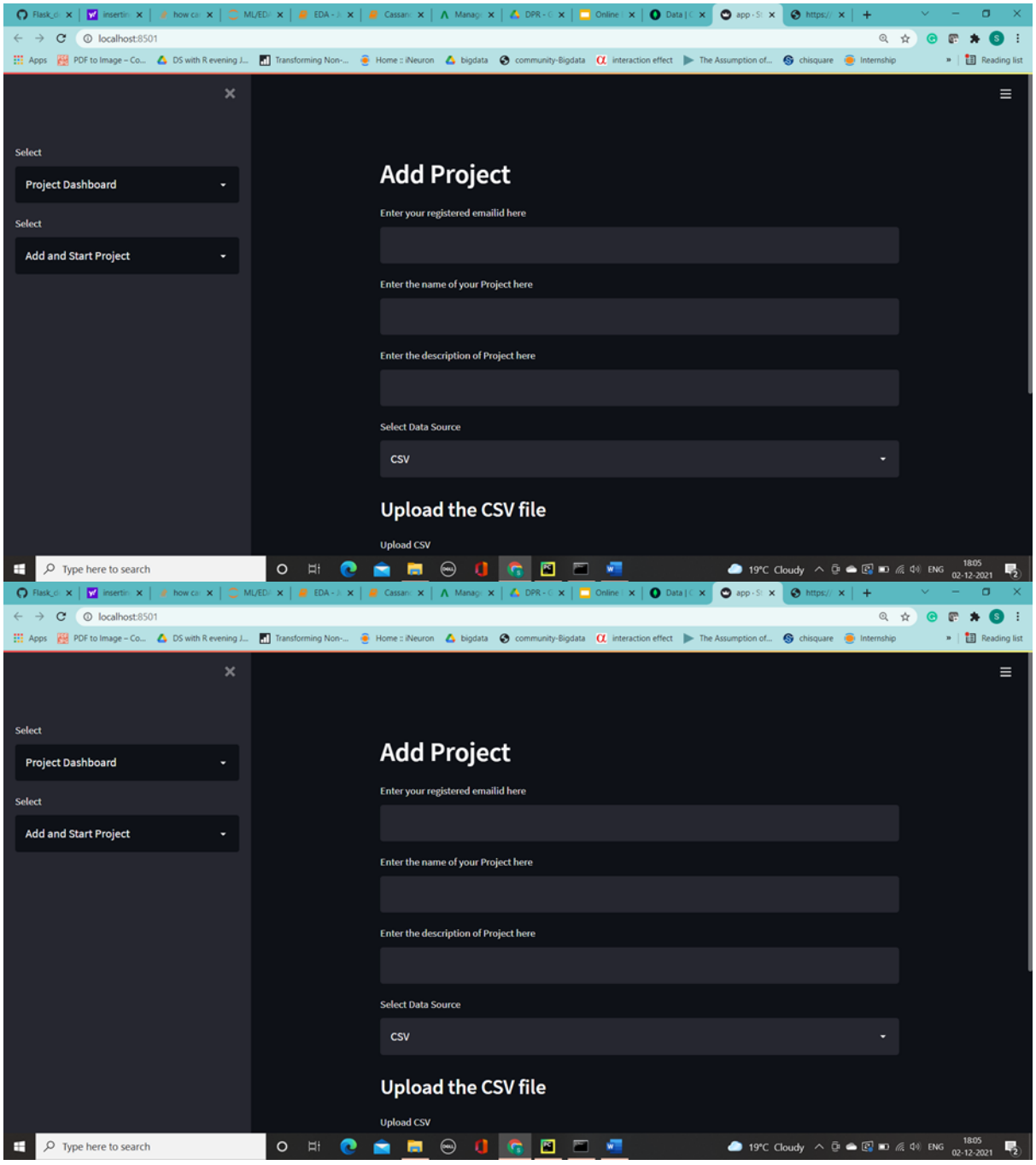


	Product_Line	Product_category	Value
0	Sporting Goods	Outdoor Recreation	5642
1	Food	Beverages & Tobacco	2514
2	Apparel & Accessories	Clothing	2365
3	Apparel & Accessories	Costumes & Accessories	4487
4	Sporting Goods	Athletics	812
5	Health & Beauty	Personal Care	6912
6	Arts & Entertainment	Hobbies & Creative Arts	5201
7	Arts & Entertainment	Paintings	8451

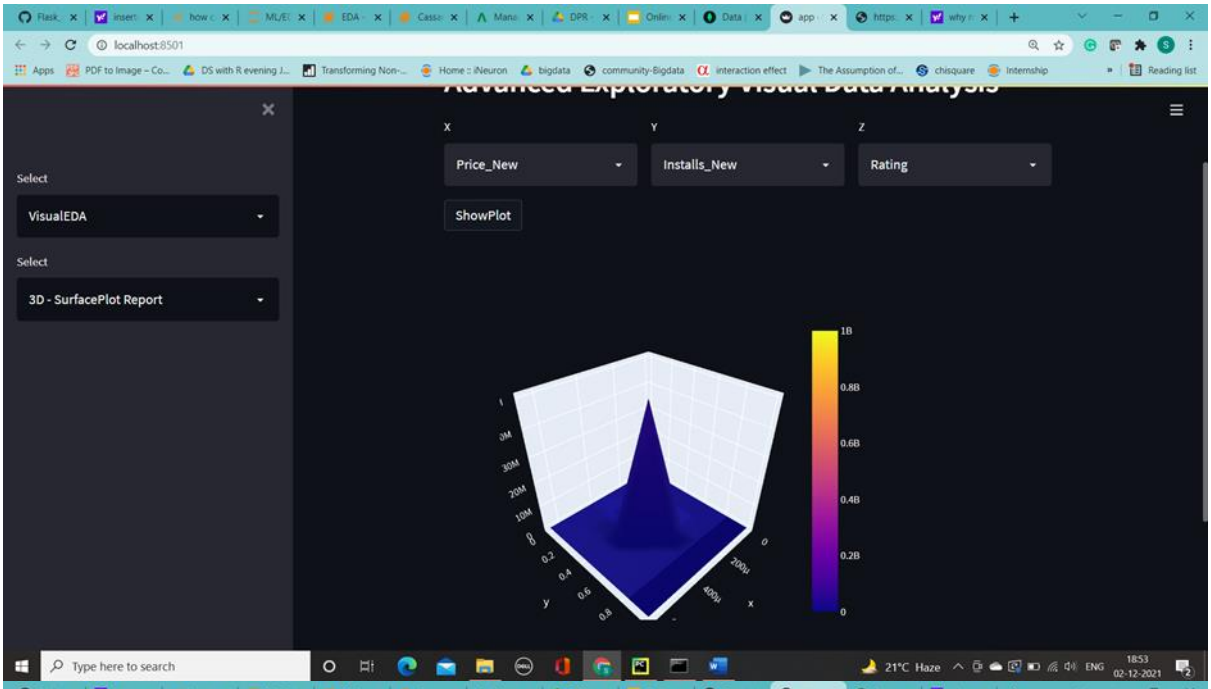

localhost:8501
Select
BasicEDA
Select
ShowDataFrame
ShowDataFrame
View Top Bottom Rows
ShowBasicDetails
PandasProfiling Report
Basic Exploratory Data Analysis
Product_Line
Product_category
Value
0 Sporting Goods Outdoor Recreation 5642
1 Food Beverages & Tobacco 2514
2 Apparel & Accessories Clothing 2365
3 Apparel & Accessories Costumes & Accessories 4487
4 Sporting Goods Athletics 812
5 Health & Beauty Personal Care 6912
6 Arts & Entertainment Hobbies & Creative Arts 5201
7 Arts & Entertainment Paintings 8451
8 Arts & Entertainment Musical Instruments 1245
9 Hardware Tool Accessories 456
Made with Streamlit
Type here to search
19°C Cloudy
ENG
18:11
02-12-2021
localhost:8501
Select
Project Dashboard
Project Dashboard
BasicEDA
AdvancedEDA
VisualEDA
Suggestion Page
0 Sporting Goods Outdoor Recreation 5642
1 Food Beverages & Tobacco 2514
2 Apparel & Accessories Clothing 2365
3 Apparel & Accessories Costumes & Accessories 4487
4 Sporting Goods Athletics 812
5 Health & Beauty Personal Care 6912
6 Arts & Entertainment Hobbies & Creative Arts 5201
7 Arts & Entertainment Paintings 8451
8 Arts & Entertainment Musical Instruments 1245
9 Hardware Tool Accessories 456
Saved file Products.csv in data folder.
Add Project
Project Added
Made with Streamlit
Type here to search
19°C Cloudy
ENG
18:11
02-12-2021

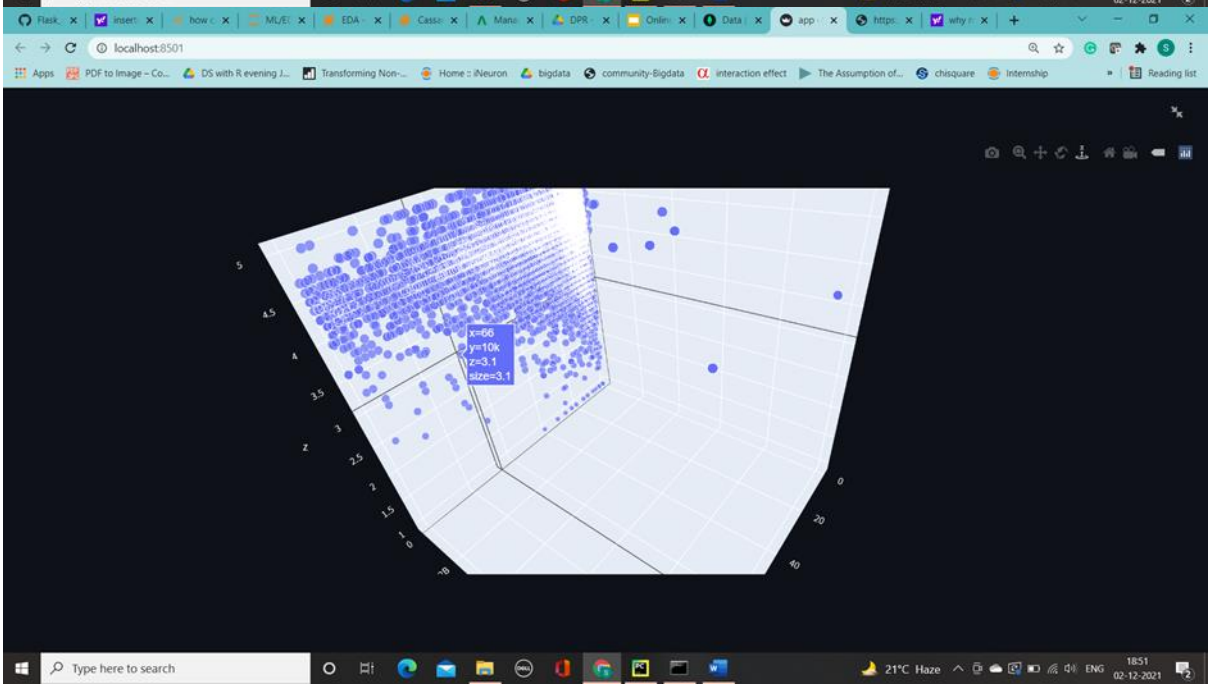


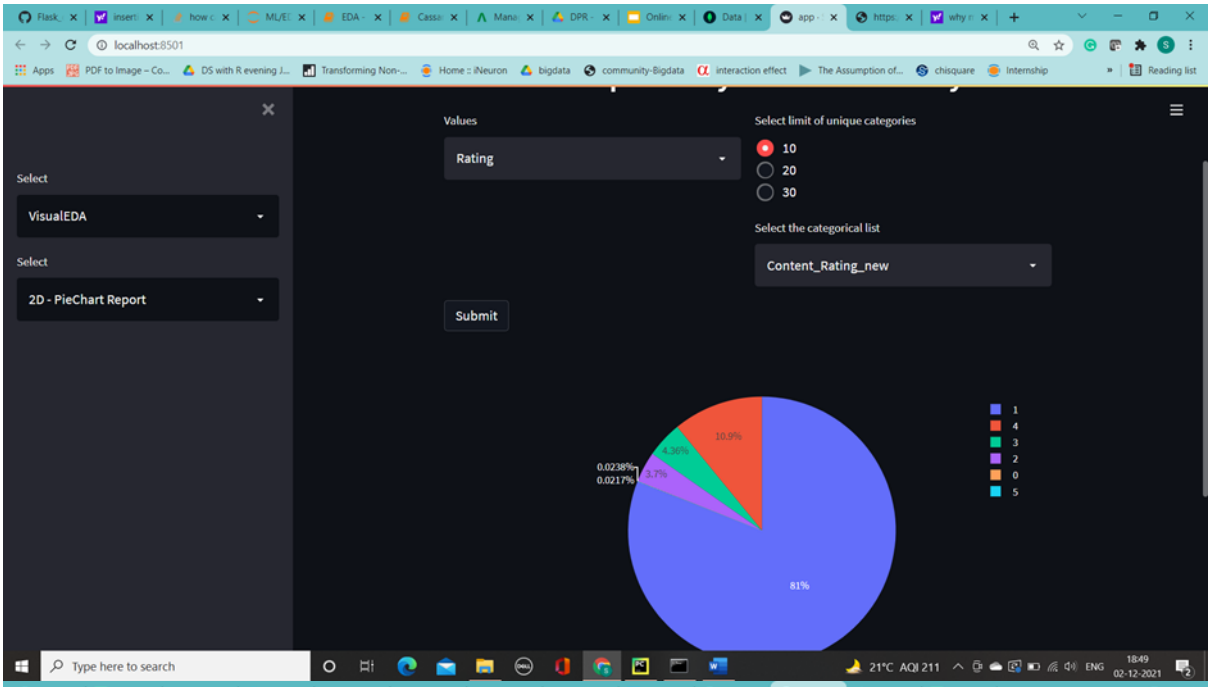
localhost:8501
Select
Project Dashboard
Select
View and Start Project
View Project
Reverify your credentials
Enter your registered emailid here
Enter password here
Enter the name of the project to work with
Submit
Type here to search
19°C Cloudy
ENG
18:07
02-12-2021

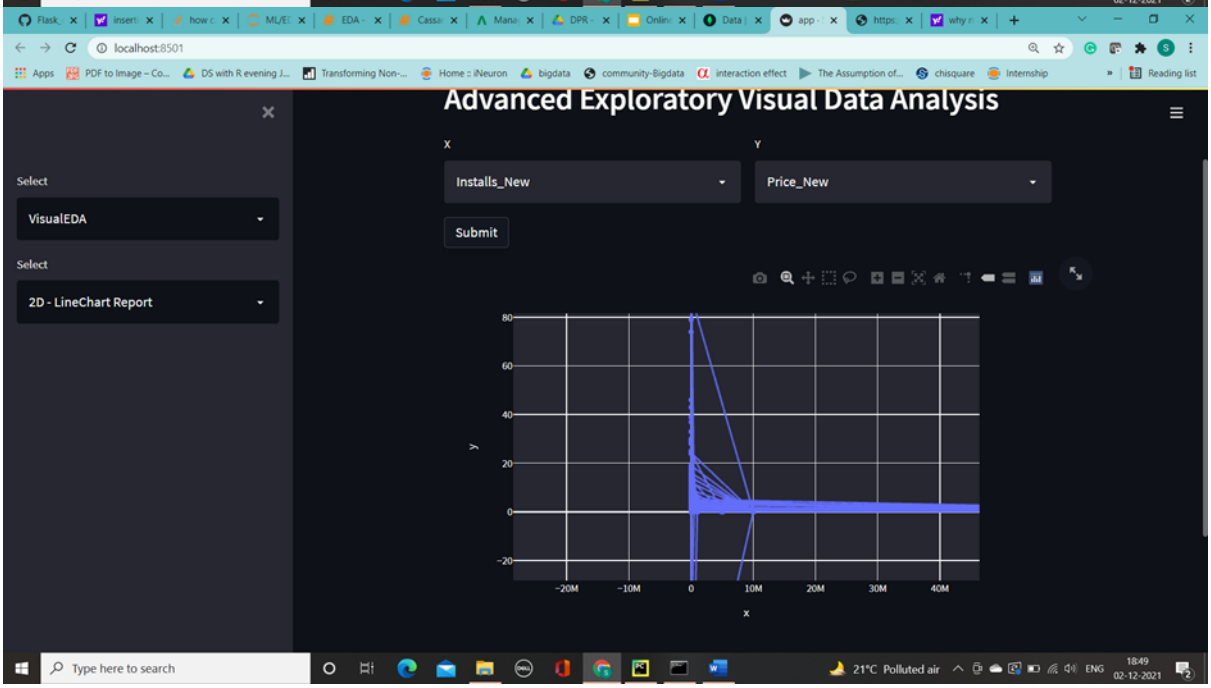


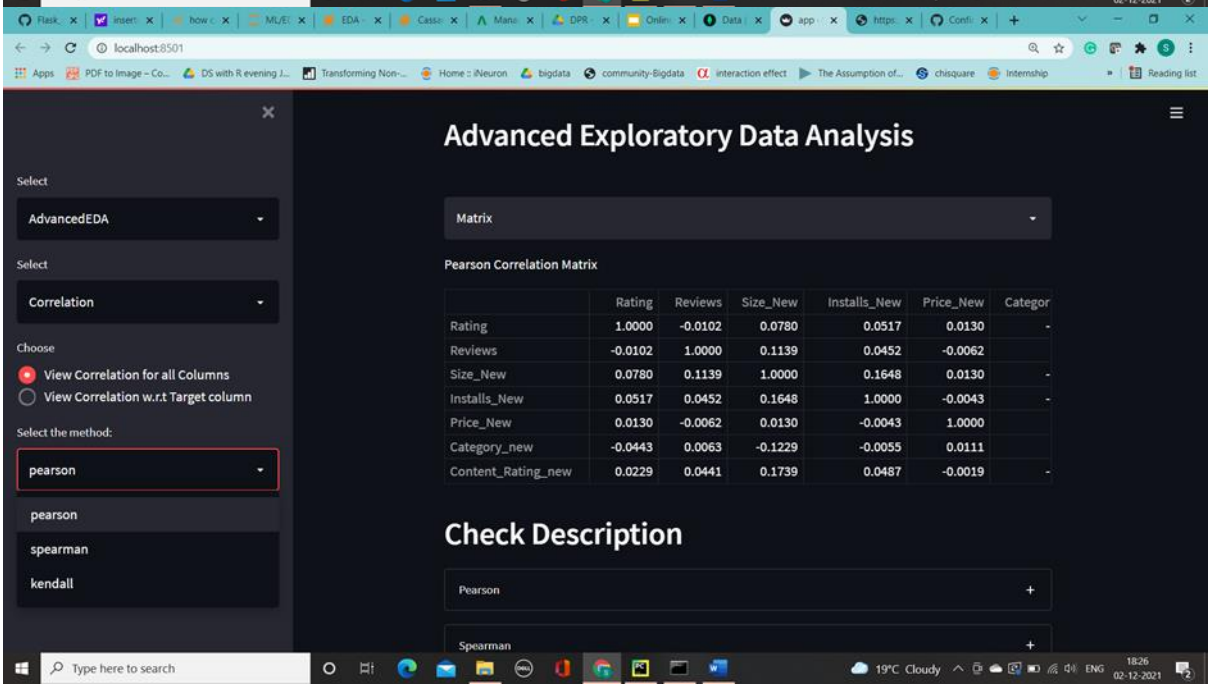
localhost:8501
Select
Project Dashboard
Select
Add and Start Project
Add Project
Enter your registered emailid here
Enter the name of your Project here
Enter the description of Project here
Select Data Source
CSV
Upload the CSV file
Upload CSV
Type here to search
19°C Cloudy
ENG
18:05
02-12-2021


localhost:8501
Select
VisualEDA
Select
3D - SurfacePlot Report
X
Price_New
Y
Installs_New
Z
Rating
ShowPlot
1B
0.8B
0.6B
0.4B
0.2B
0
Type here to search
21°C Haze
ENG
18:53
02-12-2021

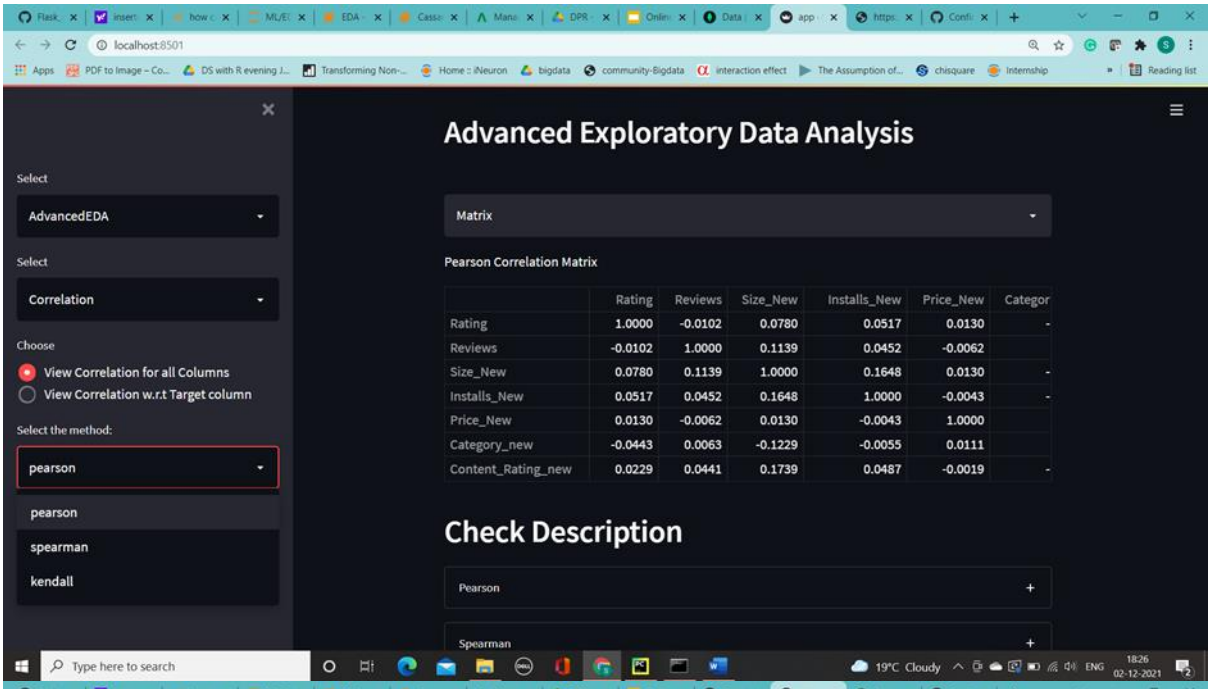

localhost:8501
x=66
y=10k
z=3.1
size=3.1
Type here to search
21°C Haze
ENG
18:51
02-12-2021

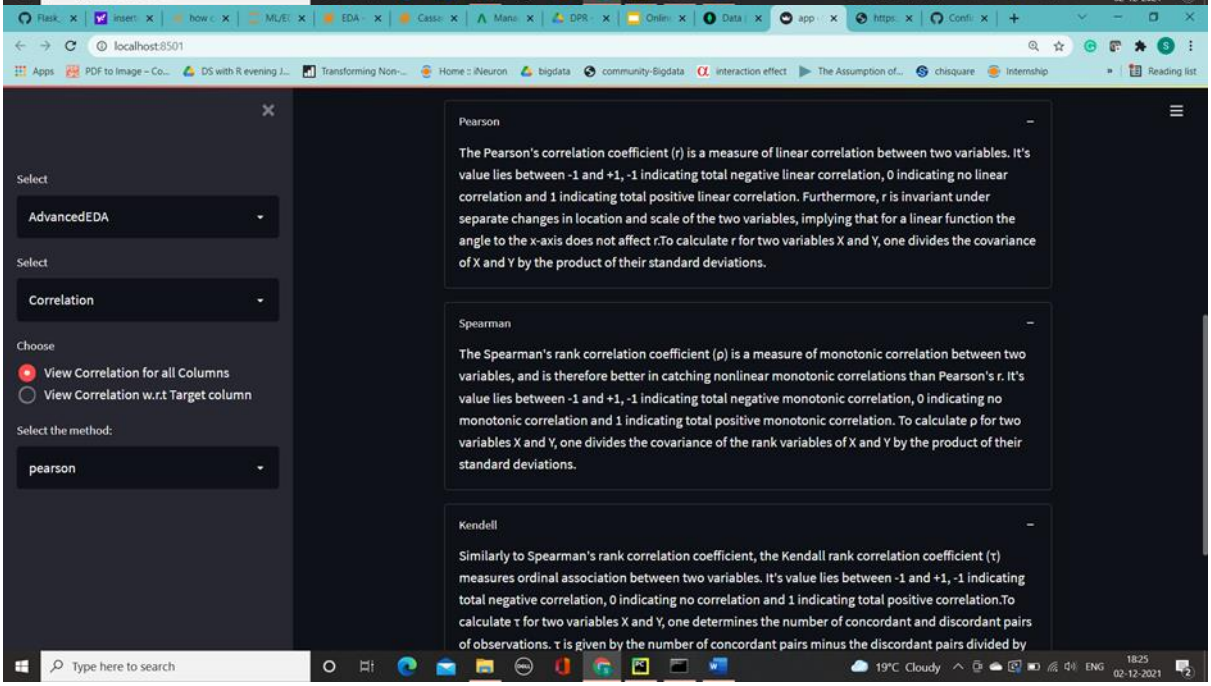

localhost:8501
Select
VisualEDA
Select
2D - PieChart Report
Values
Rating
Select limit of unique categories
10
20
30
Select the categorical list
Content_Rating_new
Submit
10.9%
4.36%
3.7%
0.0238%
0.0217%
81%
1
4
3
2
0
5
Type here to search
21°C AQI 211
ENG
18:49
02-12-2021

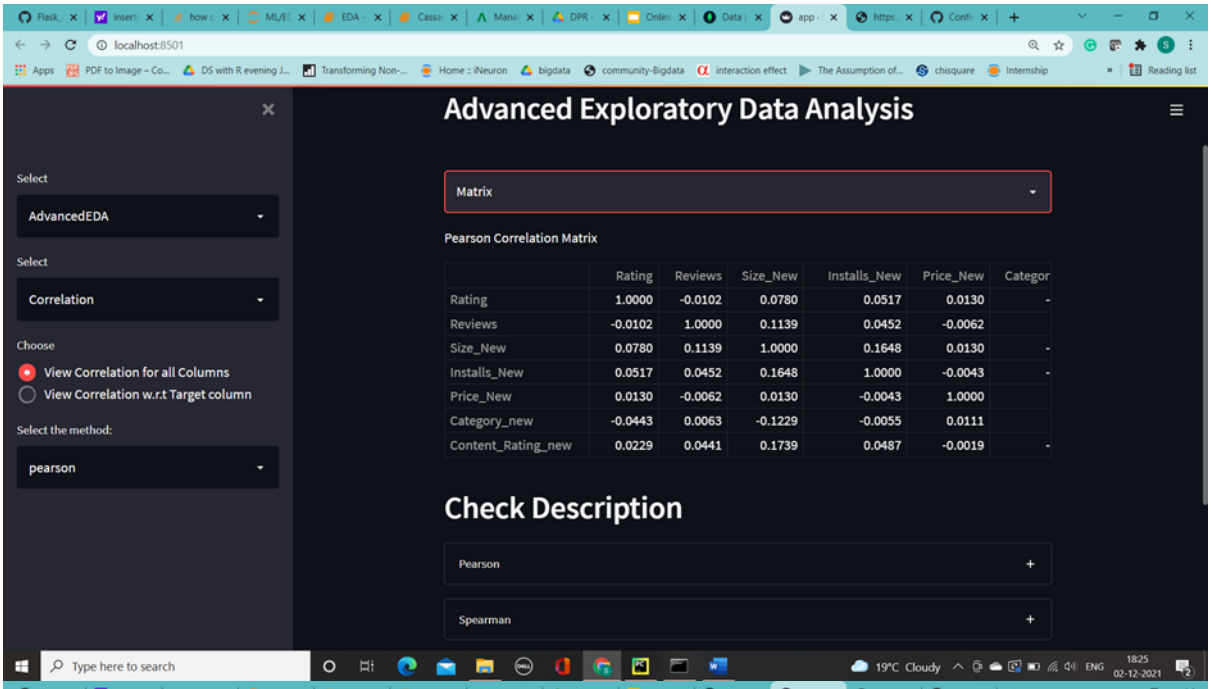

localhost:8501
Advanced Exploratory Visual Data Analysis
Select
VisualEDA
Select
2D - LineChart Report
X
Installs_New
Y
Price_New
Submit
80
60
40
20
0
-20
y
-20M
-10M
0
10M
20M
30M
40M
x
Type here to search
21°C Polluted air
ENG
18:49
02-12-2021



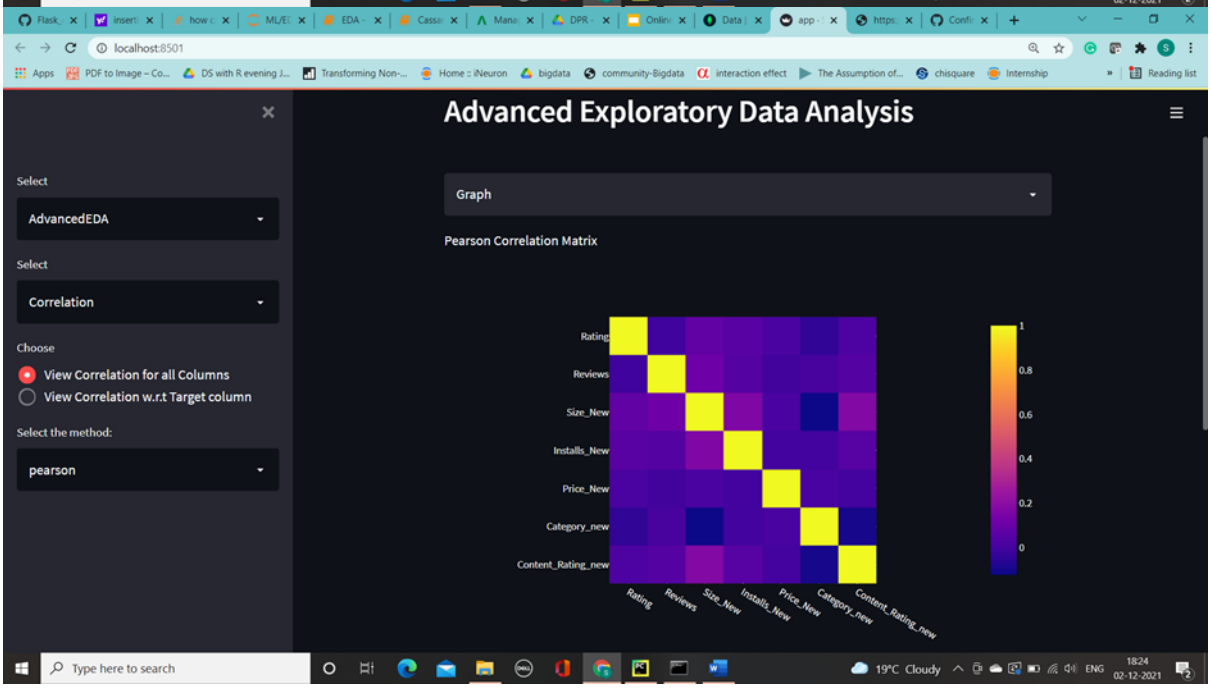
	Rating	Reviews	Size_New	Installs_New	Price_New	Categor
Rating	1.0000	-0.0102	0.0780	0.0517	0.0130	-
Reviews	-0.0102	1.0000	0.1139	0.0452	-0.0062	
Size_New	0.0780	0.1139	1.0000	0.1648	0.0130	-
Installs_New	0.0517	0.0452	0.1648	1.0000	-0.0043	-
Price_New	0.0130	-0.0062	0.0130	-0.0043	1.0000	
Category_new	-0.0443	0.0063	-0.1229	-0.0055	0.0111	
Content_Rating_new	0.0229	0.0441	0.1739	0.0487	-0.0019	-

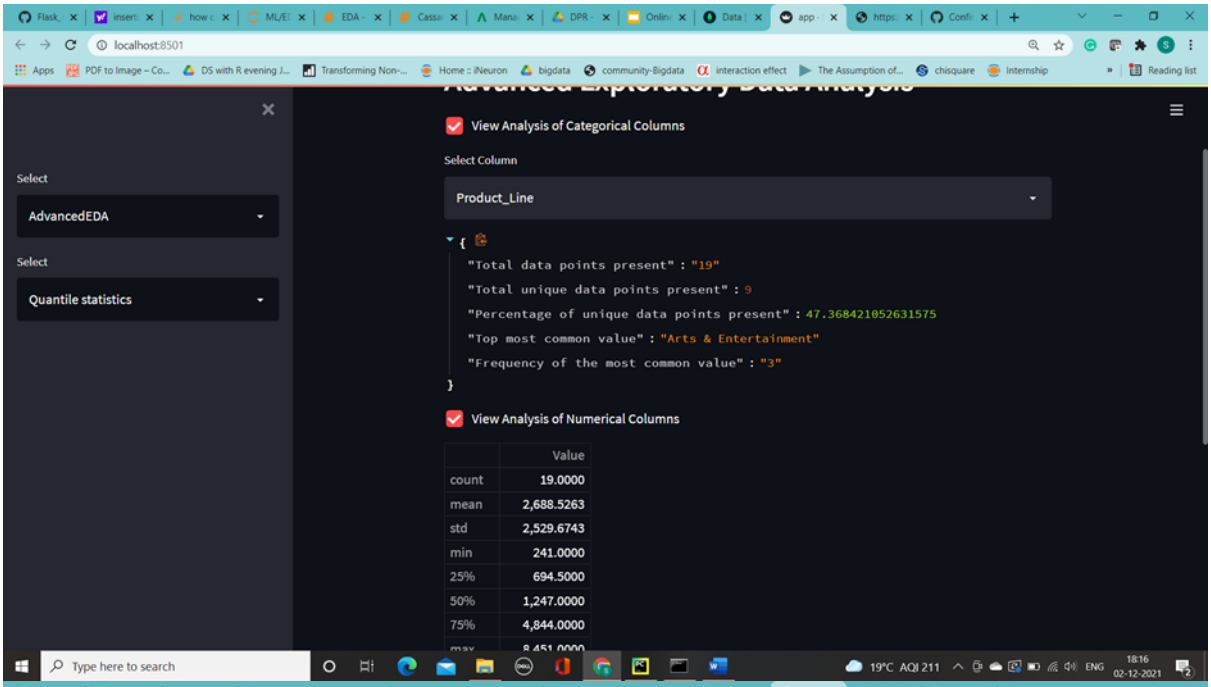

localhost:8501
Advanced Exploratory Data Analysis
Select
AdvancedEDA
Select
Correlation
Choose
View Correlation for all Columns
View Correlation w.r.t Target column
Select the method:
pearson
pearson
spearman
kendall
Matrix
Pearson Correlation Matrix
Rating Reviews Size_New Installs_New Price_New Categor
Rating 1.0000 -0.0102 0.0780 0.0517 0.0130
Reviews -0.0102 1.0000 0.1139 0.0452 -0.0062
Size_New 0.0780 0.1139 1.0000 0.1648 0.0130
Installs_New 0.0517 0.0452 0.1648 1.0000 -0.0043
Price_New 0.0130 -0.0062 0.0130 -0.0043 1.0000
Category_new -0.0443 0.0063 -0.1229 -0.0055 0.0111
Content_Rating_new 0.0229 0.0441 0.1739 0.0487 -0.0019
Check Description
Pearson
Spearman
Type here to search
19°C Cloudy
ENG
18:26
02-12-2021

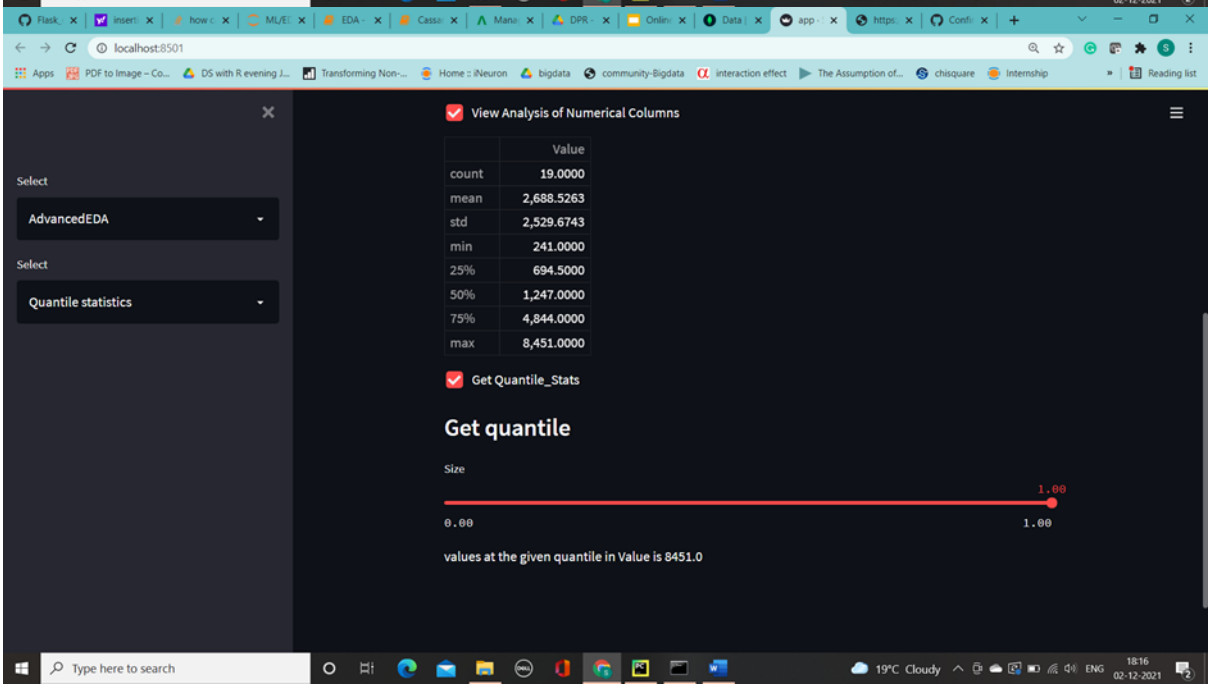

localhost:8501
Select
AdvancedEDA
Select
Correlation
Choose
View Correlation for all Columns
View Correlation w.r.t Target column
Select the method:
pearson
Pearson
The Pearson's correlation coefficient (r) is a measure of linear correlation between two variables. It's value lies between -1 and +1, -1 indicating total negative linear correlation, 0 indicating no linear correlation and 1 indicating total positive linear correlation. Furthermore, r is invariant under separate changes in location and scale of the two variables, implying that for a linear function the angle to the x-axis does not affect r.To calculate r for two variables X and Y, one divides the covariance of X and Y by the product of their standard deviations.
Spearman
The Spearman's rank correlation coefficient (ρ) is a measure of monotonic correlation between two variables, and is therefore better in catching nonlinear monotonic correlations than Pearson's r. It's value lies between -1 and +1, -1 indicating total negative monotonic correlation, 0 indicating no monotonic correlation and 1 indicating total positive monotonic correlation. To calculate ρ for two variables X and Y, one divides the covariance of the rank variables of X and Y by the product of their standard deviations.
Kendell
Similarly to Spearman's rank correlation coefficient, the Kendall rank correlation coefficient (τ) measures ordinal association between two variables. It's value lies between -1 and +1, -1 indicating total negative correlation, 0 indicating no correlation and 1 indicating total positive correlation.To calculate τ for two variables X and Y, one determines the number of concordant and discordant pairs of observations. τ is given by the number of concordant pairs minus the discordant pairs divided by
Type here to search
19°C Cloudy
ENG
18:25
02-12-2021

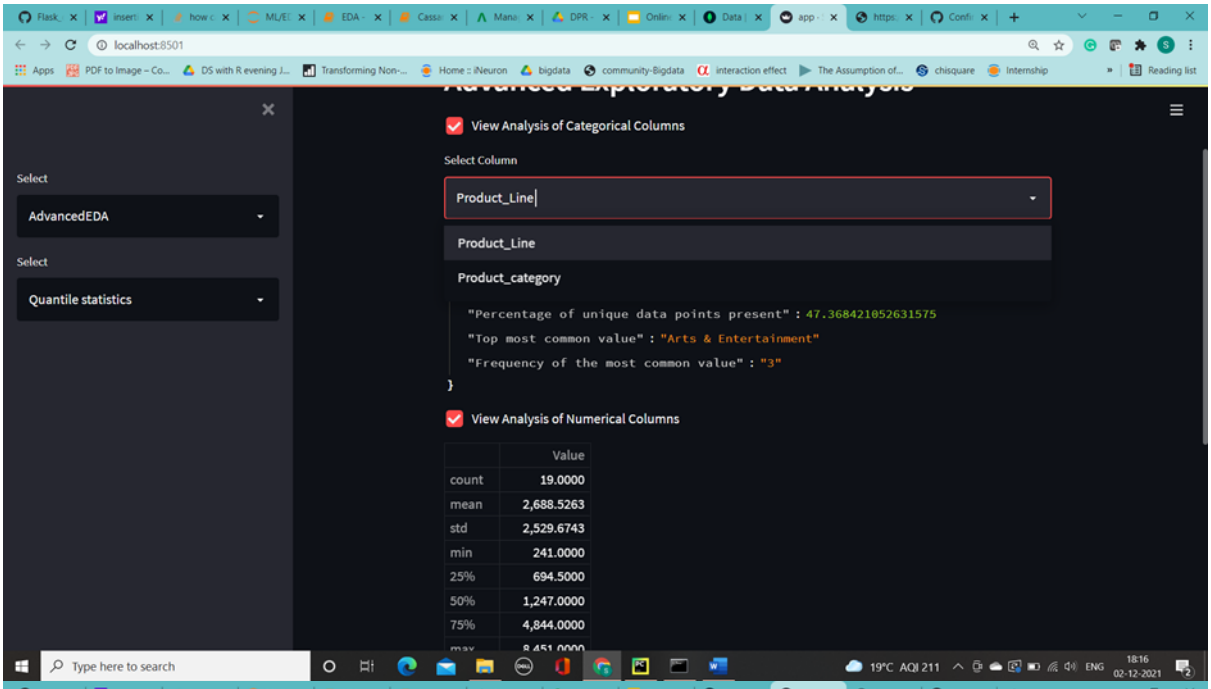


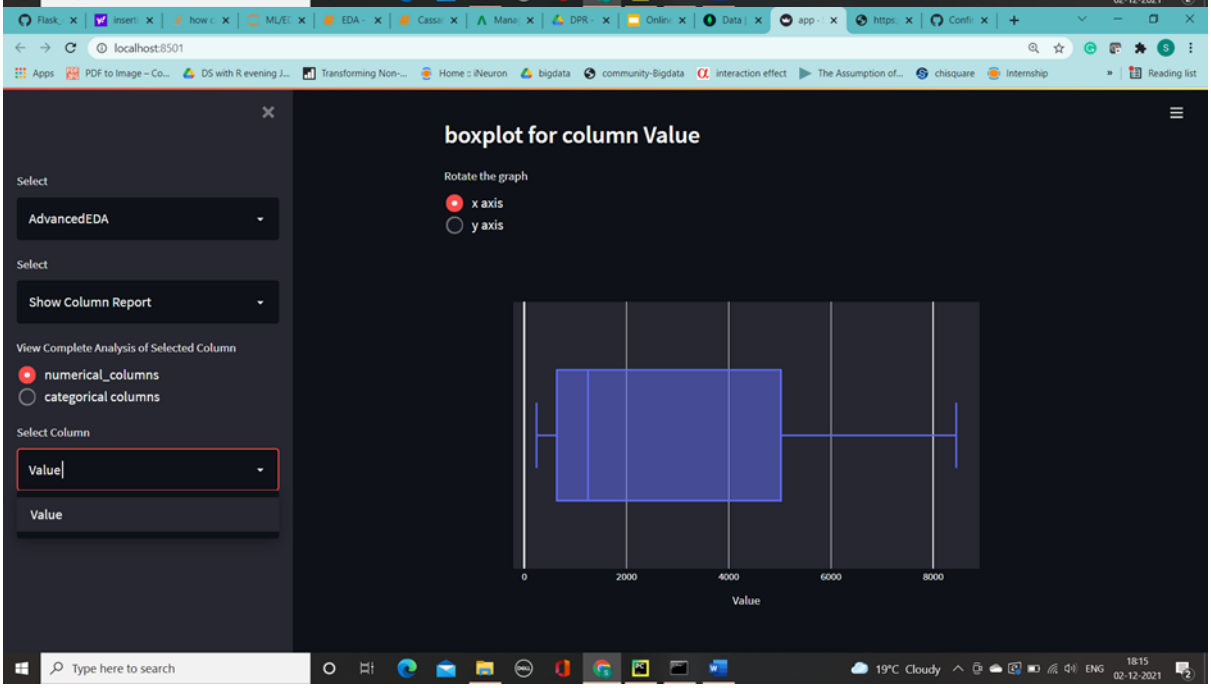
	Rating	Reviews	Size_New	Installs_New	Price_New	Categor
Rating	1.0000	-0.0102	0.0780	0.0517	0.0130	-
Reviews	-0.0102	1.0000	0.1139	0.0452	-0.0062	
Size_New	0.0780	0.1139	1.0000	0.1648	0.0130	-
Installs_New	0.0517	0.0452	0.1648	1.0000	-0.0043	-
Price_New	0.0130	-0.0062	0.0130	-0.0043	1.0000	
Category_new	-0.0443	0.0063	-0.1229	-0.0055	0.0111	
Content_Rating_new	0.0229	0.0441	0.1739	0.0487	-0.0019	-

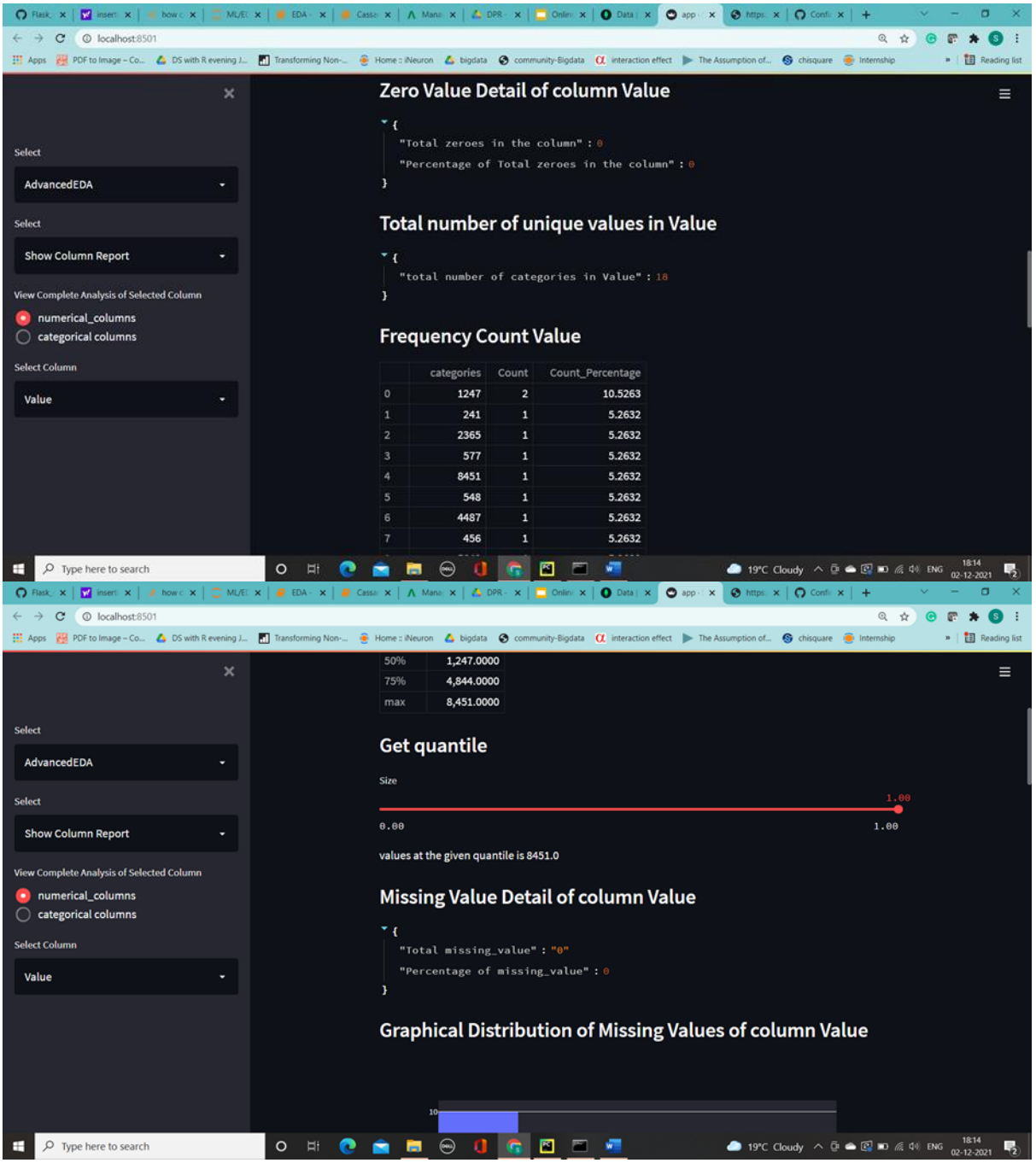



localhost:8501
Select
AdvancedEDA
Select
Quantile statistics
View Analysis of Categorical Columns
Select Column
Product_Line
{
"Total data points present" : "19"
"Total unique data points present" : 9
"Percentage of unique data points present" : 47.368421052631575
"Top most common value" : "Arts & Entertainment"
"Frequency of the most common value" : "3"
}
View Analysis of Numerical Columns
Value
count 19.0000
mean 2,688.5263
std 2,529.6743
min 241.0000
25% 694.5000
50% 1,247.0000
75% 4,844.0000
Type here to search
19°C AQI 211
ENG
18:16
02-12-2021

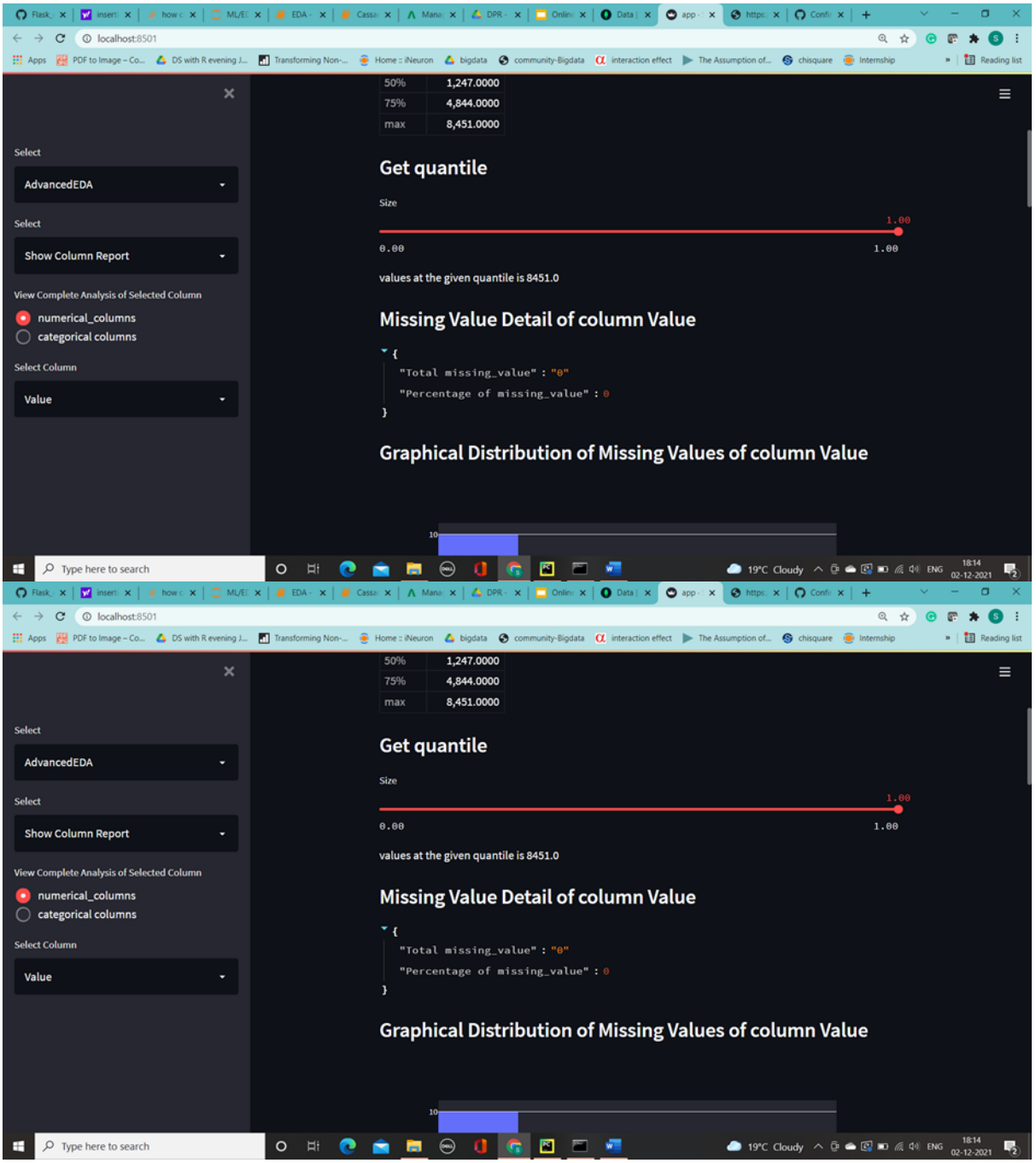

localhost:8501
Select
AdvancedEDA
Select
Quantile statistics
View Analysis of Numerical Columns
Value
count 19.0000
mean 2,688.5263
std 2,529.6743
min 241.0000
25% 694.5000
50% 1,247.0000
75% 4,844.0000
max 8,451.0000
Get Quantile_Stats
Get quantile
Size
1.00
0.00
1.00
values at the given quantile in Value is 8451.0
Type here to search
19°C Cloudy
ENG
18:16
02-12-2021

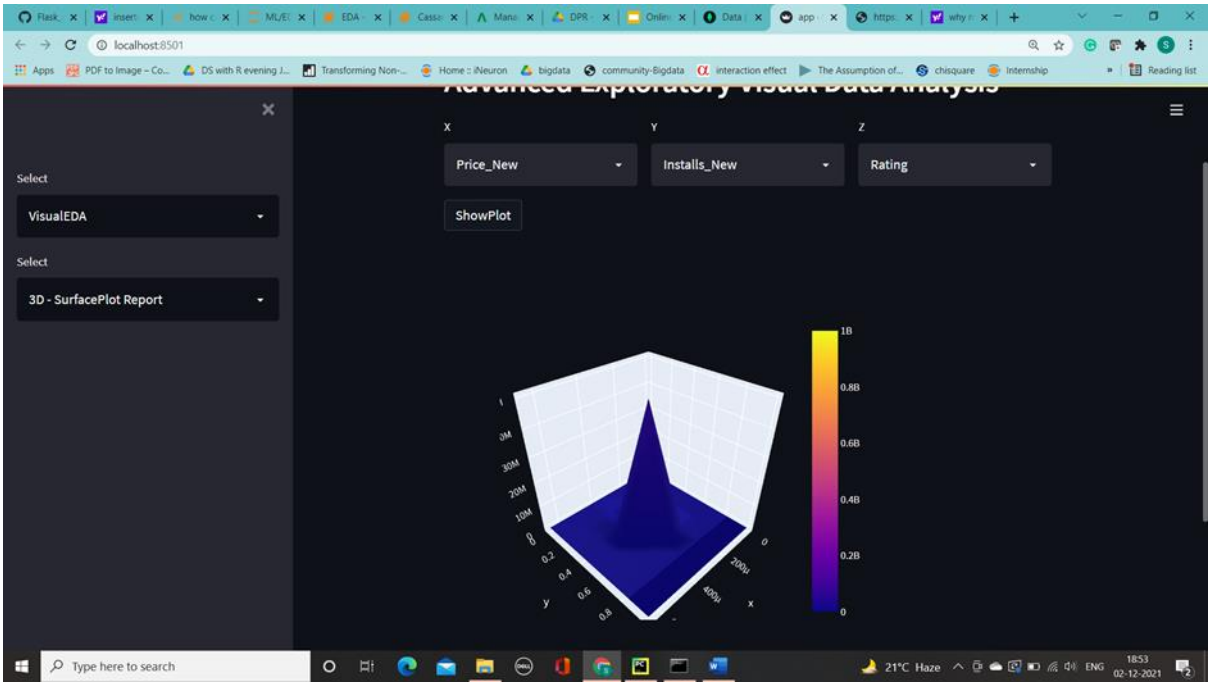

localhost:8501
Select
AdvancedEDA
Select
Quantile statistics
View Analysis of Categorical Columns
Select Column
Product_Line
Product_Line
Product_category
"Percentage of unique data points present" : 47.368421052631575
"Top most common value" : "Arts & Entertainment"
"Frequency of the most common value" : "3"
}
View Analysis of Numerical Columns
Value
count 19.0000
mean 2,688.5263
std 2,529.6743
min 241.0000
25% 694.5000
50% 1,247.0000
75% 4,844.0000
Type here to search
19°C AQI 211
ENG
18:16
02-12-2021


localhost:8501
Select
AdvancedEDA
Select
Show Column Report
View Complete Analysis of Selected Column
numerical_columns
categorical columns
Select Column
Value
Value
boxplot for column Value
Rotate the graph
x axis
y axis
0
2000
4000
6000
8000
Value
Type here to search
19°C Cloudy
ENG
18:15
02-12-2021


localhost:8501
Select
AdvancedEDA
Select
Show Column Report
View Complete Analysis of Selected Column
numerical_columns
categorical columns
Select Column
Value
Zero Value Detail of column Value
{
"Total zeroes in the column" : 0
"Percentage of Total zeroes in the column" : 0
}
Total number of unique values in Value
{
"total number of categories in Value" : 18
}
Frequency Count Value
categories Count Count_Percentage
0 1247 2 10.5263
1 241 1 5.2632
2 2365 1 5.2632
3 577 1 5.2632
4 8451 1 5.2632
5 548 1 5.2632
6 4487 1 5.2632
7 456 1 5.2632
Type here to search
19°C Cloudy
ENG
18:14
02-12-2021

localhost:8501
50% 1,247.0000
75% 4,844.0000
max 8,451.0000
Select
AdvancedEDA
Select
Show Column Report
View Complete Analysis of Selected Column
numerical_columns
categorical columns
Select Column
Value
Get quantile
Size
1.00
0.00
1.00
values at the given quantile is 8451.0
Missing Value Detail of column Value
{
"Total missing_value" : "0"
"Percentage of missing_value" : 0
}
Graphical Distribution of Missing Values of column Value
10
Type here to search
19°C Cloudy
ENG
18:14
02-12-2021


localhost:8501
50% 1,247.0000
75% 4,844.0000
max 8,451.0000
Select
AdvancedEDA
Select
Show Column Report
View Complete Analysis of Selected Column
numerical_columns
categorical columns
Select Column
Value
Get quantile
Size
1.00
0.00
1.00
values at the given quantile is 8451.0
Missing Value Detail of column Value
{
"Total missing_value" : "0"
"Percentage of missing_value" : 0
}
Graphical Distribution of Missing Values of column Value
10
Type here to search
19°C Cloudy
ENG
18:14
02-12-2021


localhost:8501
Select
VisualEDA
Select
3D - SurfacePlot Report
X
Price_New
Y
Installs_New
Z
Rating
ShowPlot
1B
0.8B
0.6B
0.4B
0.2B
0
Type here to search
21°C Haze
ENG
18:53
02-12-2021

-