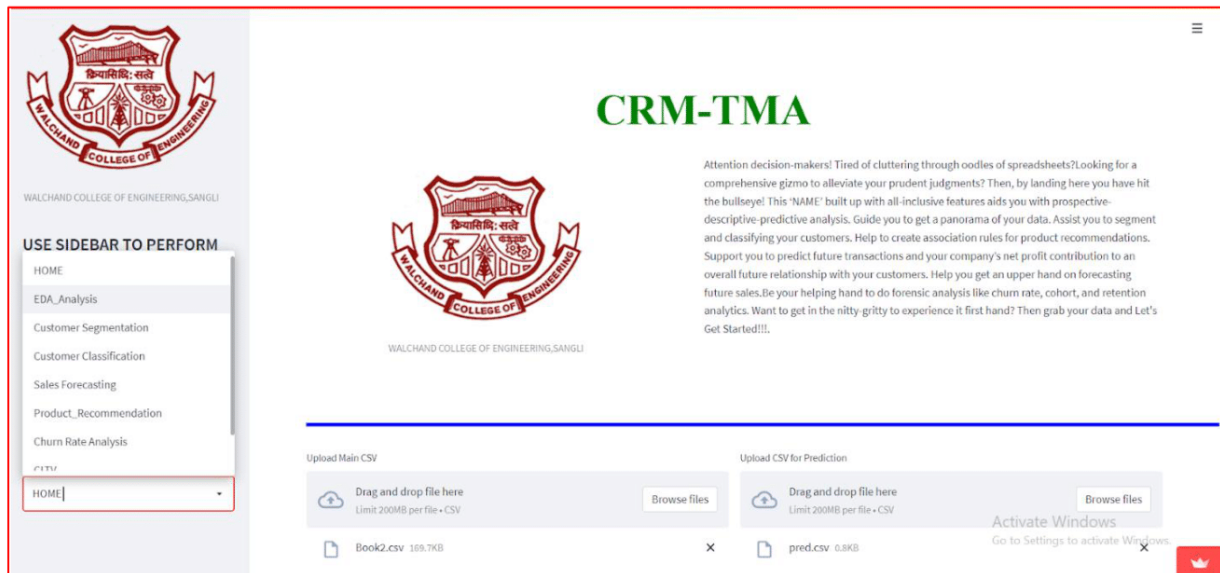


EDA ANALYSIS

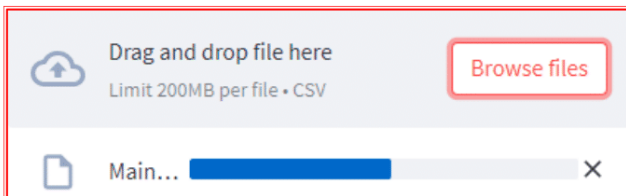
A GUIDE TO USER



2.1 EDA Analysis



- Select the EDA Analysis option from the dropdown in the sidebar menu. Wait for some time until the screen gets static.

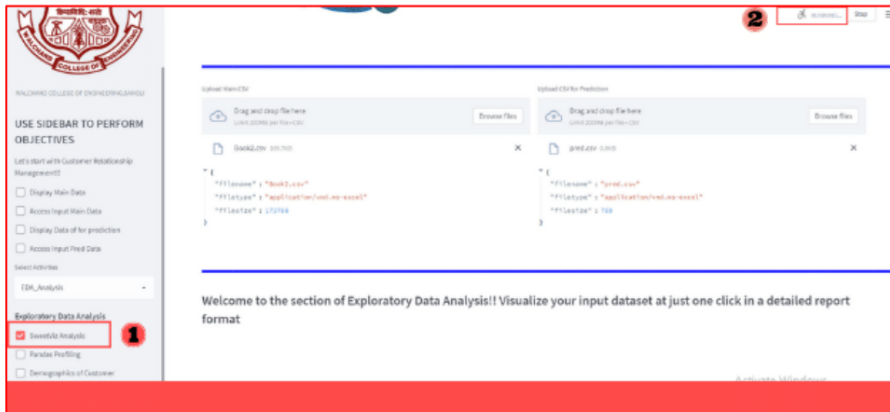


- Upload the **.csv file** of the dataset whose EDA Analysis is to be done
- This can be done either by 'drag and drop' mode or by just browsing your directory.

- Further you can choose between Sweetviz Analysis, Pandas Profiling and Demographics of Customer as per your requirements by clicking the checkbox in the sidebar menu.

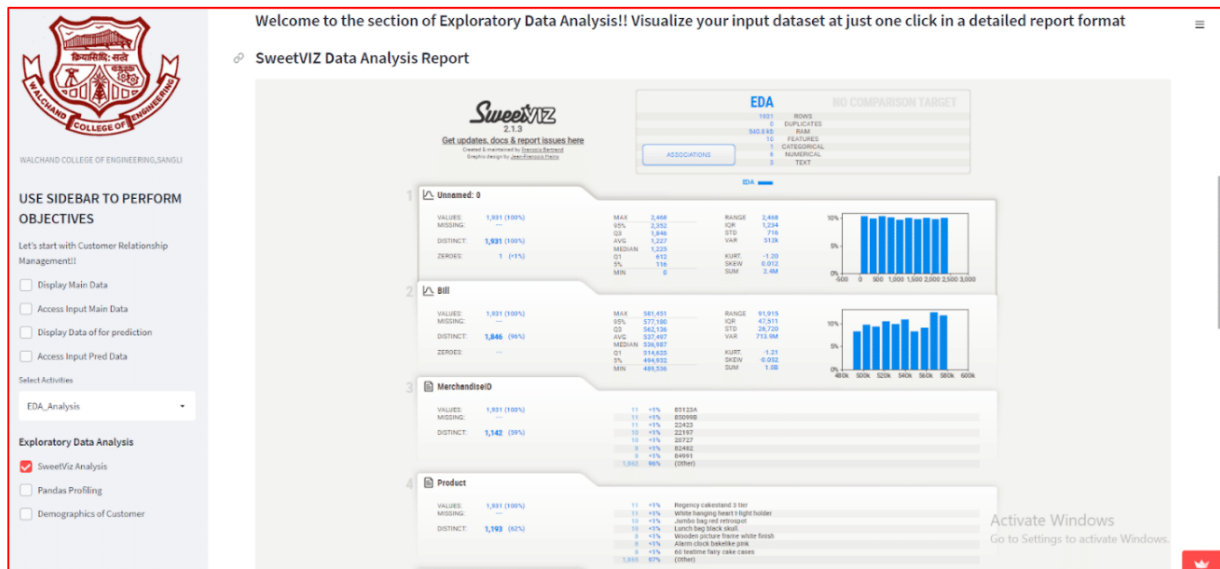


2.2 SweetViz Analysis



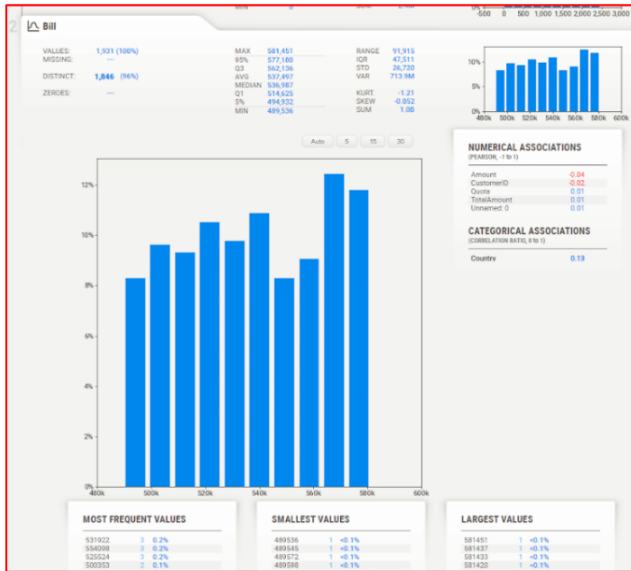
- Click on the **Sweetviz Analysis** option in the sidebar menu. Wait for some time until the results are displayed.

- Sweetviz data Analysis Report will be generated.



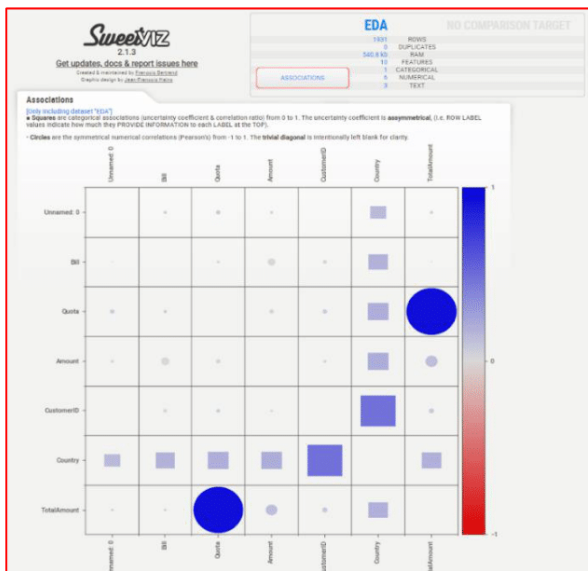
- Here you have various options to understand the crux of data. Right from visualizing the statistical information of each and every attribute to associations between attributes.
- Here by clicking on any one of the attribute you will get detail analysis.

Grahak 360 CRM User Manual



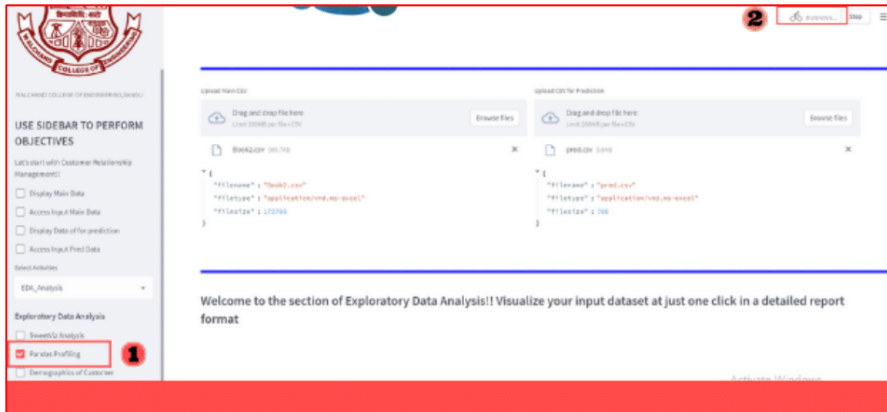
- Get insights on sparse and distinct values.
- Get numerical and categorical associations.
- Get the details of most frequent, smallest and largest values.
- Get visualization of the attribute.

- Here by clicking on Associations at the right most tab of Sweetviz Analysis you will get detail associations insights.



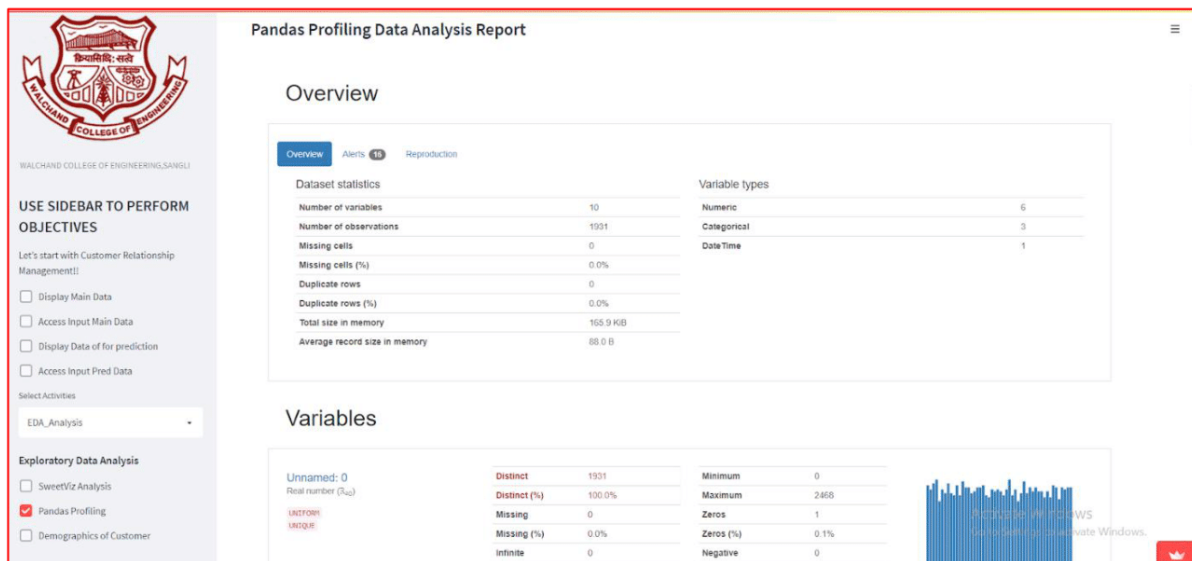
- Get insights on categorical associations represented in squares to understand correlations between attributes.
- Get insights on numerical associations represented in circles to understand correlations between attributes.

2.3 Pandas Profiling Analysis

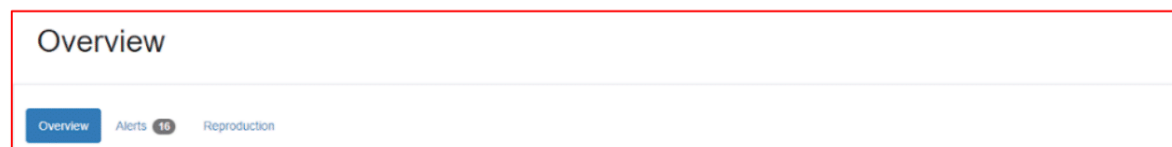


- Click on the **Pandas Profiling** option in the sidebar menu. Wait for some time until the results are displayed.

- Pandas Profiling Analysis Report will be generated.



- Here you have six different options to help visualize and analyze your dataset. Pandas profiling analysis will help you in getting an overview of dataset, analysis of variables, Interaction graphs between variables, correlation analysis, insights on missing values and display of a sample dataset.
- Here by clicking on overview option you will get three sub options which Includes Overview of dataset, Alerts and Reproduction of your analysis.



- Here by clicking on overview of dataset sub option you will get detail analysis.

Overview

Overview Alerts 15 Reproduction

Dataset statistics		Variable types	
Number of variables	10	Numeric	6
Number of observations	1931	Categorical	5
Missing cells	0	Date Time	1
Missing cells (%)	0.0%		
Duplicate rows	0		
Duplicate rows (%)	0.0%		
Total size in memory	165.9 KB		
Average record size in memory	86.0 B		

- Get insights on dataset statistics to understand number of variables, duplicate rows, sparse data etc.
- Get insights on variable types of your attribute.

- Here by clicking on Alerts sub option you will get detail analysis.

Overview

Overview Alerts 15 Reproduction

Alerts

PersonnelId(22) has a high cardinality: 1142 distinct values	High cardinality
Product() has a high cardinality: 1193 distinct values	High cardinality
Quota() is highly correlated with TotalDiscount	High correlation
TotalDiscount() is highly correlated with Quota	High correlation
Quota() is highly correlated with TotalDiscount	High correlation
TotalDiscount() is highly correlated with Quota	High correlation
Quota() is highly correlated with TotalDiscount	High correlation
TotalDiscount() is highly correlated with Quota	High correlation
Quota() is highly correlated with TotalDiscount	High correlation
Country() is highly correlated with CustomerID	High correlation
CustomerID() is highly correlated with Country	High correlation
TotalDiscount() is highly correlated with Quota	High correlation
Quota() is highly skewed (y1 = 41.40965742)	Skewed
Quota() is uniformly distributed	Uniform
Product() is uniformly distributed	Uniform
Quota() has unique values	Unique

- Get top insights on the notifications provided to understand which attributes has high cardinality, high correlation, skewed data, uniform data and unique data in one glance.

Overview

Overview Alerts 16 Reproduction

Reproduction

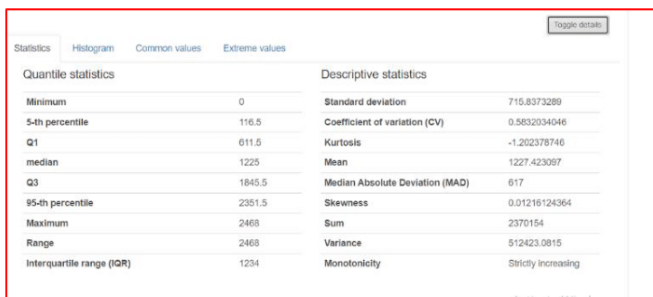
Analysis started	2022-03-13 03:18:57.845334
Analysis finished	2022-03-13 03:19:06.817617
Duration	8.97 seconds
Software version	pandas-profiling v3.1.0
Download configuration	config.json

- Here by clicking on Reproduction sub option you will get detail analysis.
- Get information on when the analysis was started, finished and other relevant metadata.

- Here by clicking on variables option you will get detail statistical analysis of every attribute and also by clicking on toggle details you will get more analysis and visualization associated with it. This toggle details further has 4 sub options: statistics, histogram, common values and extreme values visualization.

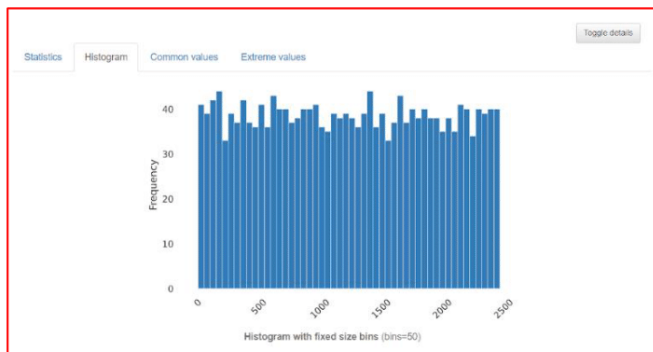


- Here by clicking on statistics sub-option you will get detail analysis.



- Get insights on Quantile as well as Descriptive Statistics.

- Here by clicking on Histogram sub-option you will get detail analysis.



- Get insights on Histogram plots.

- Here by clicking on common values sub-option you will get detail analysis.

Statistics Histogram Common values Extreme values

Value	Count	Frequency (%)
0	1	0.1%
1690	1	0.1%
1652	1	0.1%
1648	1	0.1%
1647	1	0.1%
1646	1	0.1%
1645	1	0.1%
1644	1	0.1%
1643	1	0.1%
1642	1	0.1%
Other values (1921)	1921	99.5%

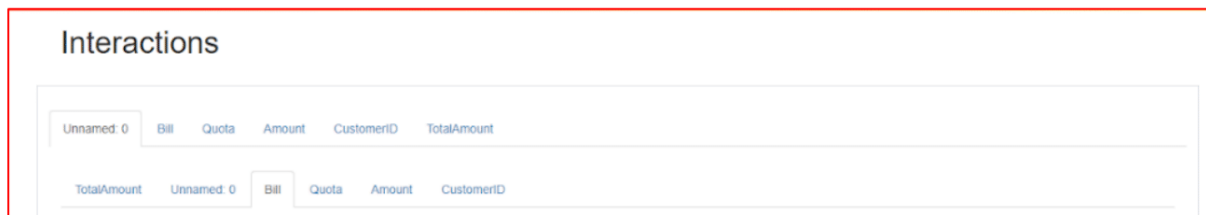
- Get insights on count and frequency of the distinct values present in the attributes.

- Here by clicking on extreme values sub-option you will get detail analysis.

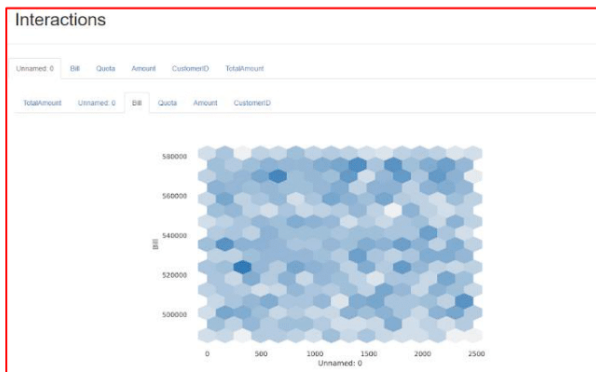
Minimum 5 values		Maximum 5 values	
Value	Count	Frequency (%)	
0	1	0.1%	
1	1	0.1%	
2	1	0.1%	
3	1	0.1%	
4	1	0.1%	
5	1	0.1%	
6	1	0.1%	
8	1	0.1%	
9	1	0.1%	
10	1	0.1%	

- Get insights on count and frequency of extreme minimum and maximum values.

- Here by clicking on Interaction option you will get set of variables to be plotted on x and y axis to understand there correlations.



- Here by choosing between variables to be plotted on x and y axis you can get a correlation plots.



- Get insights on dependency of the variables.
- Will help you find hidden patterns between the correlations.

- Here by clicking on Interactions option you will get detail analysis.
- Here by clicking on Correlations option you can select between Spearman's, Pearson's, Kendall's and Phik rank correlation coefficient as per your requirements to understand the correlations in detail. Also with help of Toggle correlation descriptions you can actually get more information about it in descriptive form.

Correlations

Spearman's ρ Pearson's r **Kendall's τ** Phik (ϕ_k)

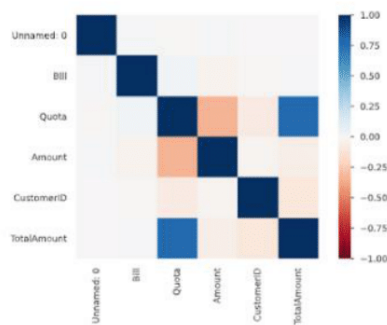
Toggle correlation descriptions

- Here by clicking on one of correlation method option you will get detail analysis of the chosen method.

Correlations

Spearman's ρ Pearson's r **Kendall's τ** Phik (ϕ_k)

Toggle correlation descriptions



Kendall's τ

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 the total number of pairs.

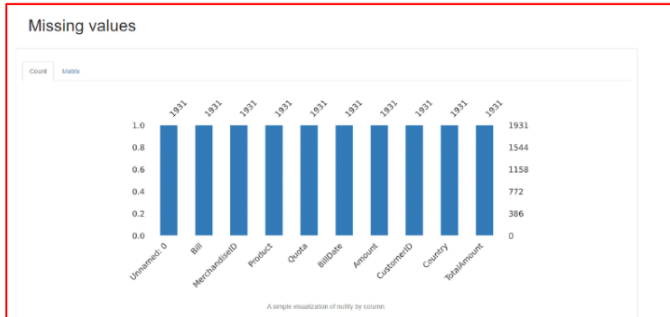
- By clicking on say Kendall's T option you will get a heat map analysis describing the correlation.
- In case want the descriptive analysis of the same then click on Toggle correlation descriptions.
- Here by clicking on the missing value you will get two options to visualize the sparse data in form of count and matrix form.

Missing values

Count **Matrix**

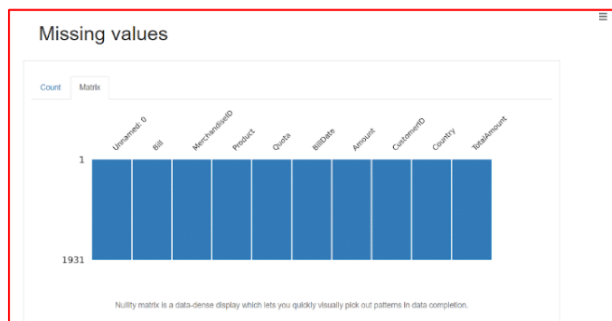
Grahak 360 CRM User Manual

- Here by clicking on the Count Option you will get visualization of the sparse data.



- Get insights on a simple visualization of nullity by columns.

- Here by clicking on the Matrix Option you will get visualization of the sparse data.



- From this data-dense null matrix get quicker insights on visualizing patterns in data completion.

- Here by clicking on Sample option get quicker insights on visualizing the first and last rows of your dataset.

Sample

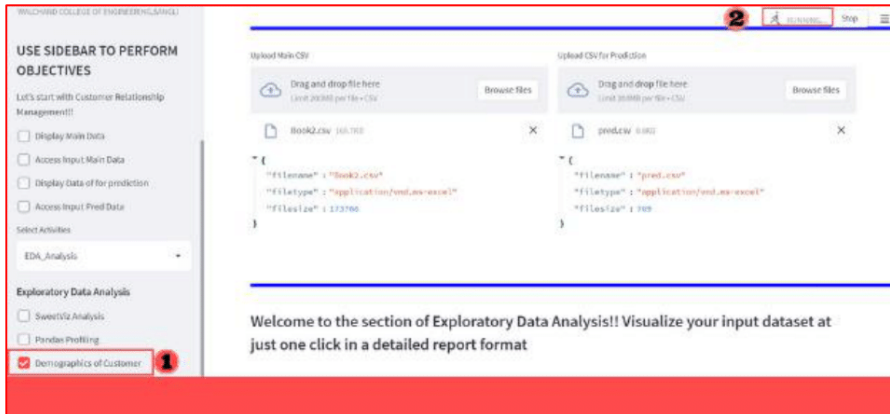
First rows

Unnamed: 0	Bill	MerchandiseID	Product	Quota	BillDate	Amount	CustomerID	Country	TotalAmount	
0	0	532657	21314	Small glass heart trinket pot	12	2018-11-14	4.10	14562	United Kingdom	49.20
1	1	563214	22383	Lunch bag suki design	2	2019-08-14	3.65	16370	United Kingdom	7.30
2	2	507597	22561	Wooden school colouring set	12	2018-10-05	3.65	17700	United Kingdom	43.80
3	3	491634	21588	Retro spot giant tube matches	1	2017-11-12	4.55	17841	United Kingdom	4.55
4	4	496007	85232B	Set/3 russian doll stacking tins	3	2018-01-28	6.95	15203	United Kingdom	20.85
5	5	539041	21832	Chocolate calculator	4	2018-12-15	3.65	15456	United Kingdom	14.60
6	6	575905	22089	Paper bunting vintage paisley	6	2019-11-11	4.95	13732	United Kingdom	29.70
7	8	536804	22988	Soldiers egg cup	72	2018-02-12	3.25	14031	United Kingdom	234.00
8	9	546899	20719	Woodland charlotte bag	50	2019-03-17	2.72	14298	United Kingdom	136.00
9	10	514369	22396	Door mat airmail	2	2018-02-07	9.49	12477	Germany	18.98

Last rows

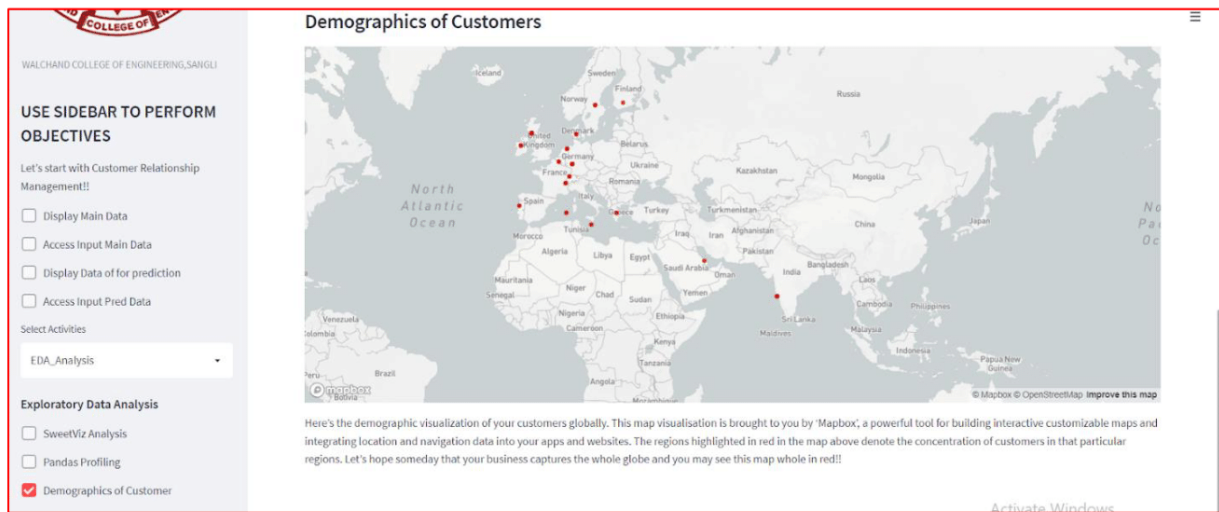
Unnamed: 0	Bill	MerchandiseID	Product	Quota	BillDate	Amount	CustomerID	Country	TotalAmount	
1921	2457	511307	21653	Glass etched t-light holder large	2	2018-07-06	4.10	13983	United Kingdom	8.20
1922	2458	512948	71477	Colour glass star t-light holder	3	2018-06-20	5.25	13230	United Kingdom	15.75
1923	2460	579267	23206	Jumbo bag pears	10	2019-11-29	4.08	14961	United Kingdom	40.80
1924	2461	501031	21488	Red white scarf hot water bottle	4	2018-11-03	5.95	16718	United Kingdom	23.80
1925	2462	531728	22424	Enamel bread bin cream	2	2018-09-11	14.75	16223	United Kingdom	29.50
1926	2463	530843	22764	Rustic wooden cabinet, glass doors	2	2018-04-11	26.95	14895	United Kingdom	53.90
1927	2464	513975	22321	Bird decoration red spot	9	2018-06-29	2.85	13451	United Kingdom	25.65
1928	2465	580541	23166	Medium ceramic tea storage jar	12	2019-05-12	3.25	15358	United Kingdom	39.00

2.4 Demographics of Customer



- Click on the **Demographics of Customer** option in the sidebar menu. Wait for some time until the results are displayed.

- Sweetviz data Analysis Report will be generated.



- Here you can visualize demographic spread of your customers globally.
- To fully view the demographics of customers, use the expander at the top right corner
- The regions highlighted in red in the map helps you to visualize the concentration of the customers in particular regions.