

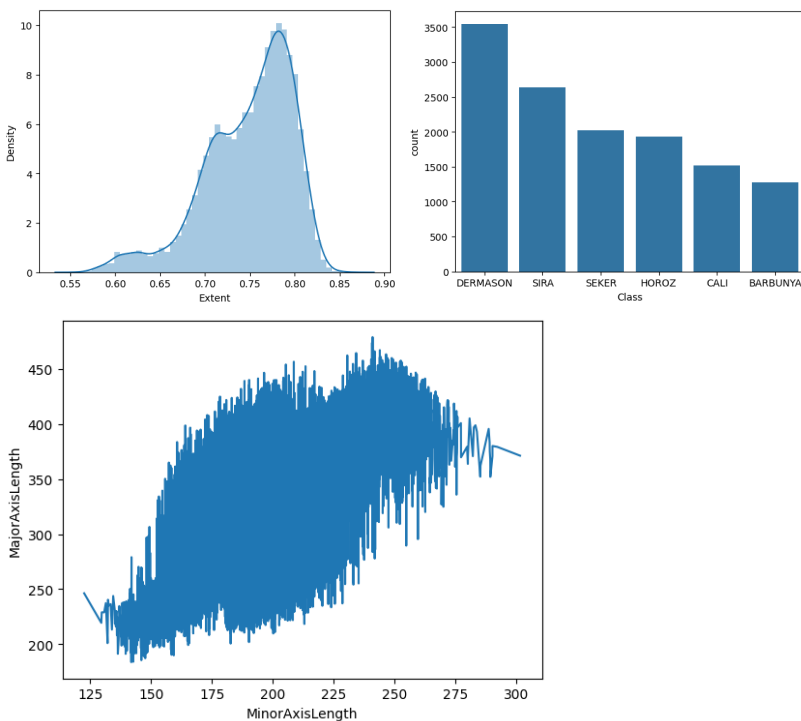
Data Collection and Preprocessing Phase

Date	10 July 2024
Team ID	739713
Project Title	Beansense: Precision Bean Classification For Enhanced Agricultural And Culinary Applications
Maximum Marks	6 Marks

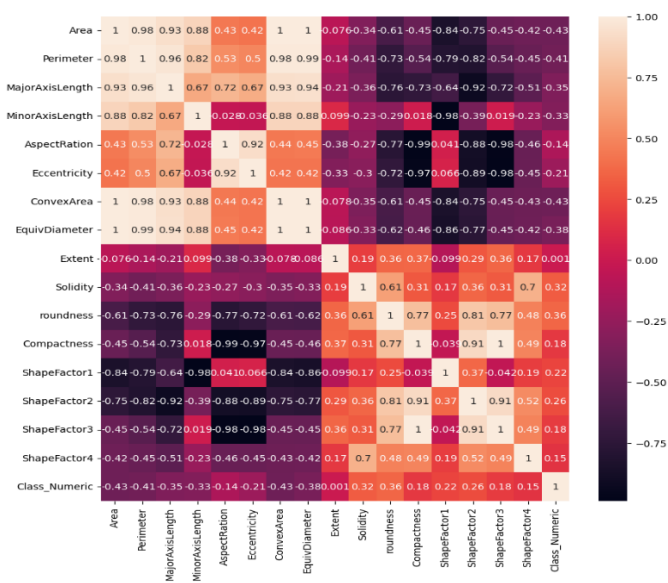
Data Exploration and Preprocessing Report

Dataset variables will be statistically analyzed to identify patterns and outliers, with Python employed for preprocessing tasks like normalization and feature engineering. Data cleaning will address missing values and outliers, ensuring quality for subsequent analysis and modeling, and forming a strong foundation for insights and predictions.

Section	Description																																																																																																																																																																																																																																				
Data Overview	<div>Dimension:</div> <div>13611 rows × 18 columns</div> <div>Descriptive statistics:</div> <table><thead><tr><th></th><th>Area</th><th>Perimeter</th><th>MajorAxisLength</th><th>MinorAxisLength</th><th>AspectRatio</th><th>Eccentricity</th><th>ConvexArea</th><th>EquivalentDiameter</th><th>Extent</th><th>Solidity</th><th>roundness</th><th>Compactness</th><th>Shapefactor1</th><th>Shapefactor2</th><th>Shapefactor3</th><th>Shapefactor4</th><th>Class</th><th>Class</th></tr></thead><tbody><tr><td>0</td><td>28395</td><td>610.291</td><td>208.178117</td><td>173.888747</td><td>1.197191</td><td>0.548812</td><td>28715</td><td>190.141097</td><td>0.763923</td><td>0.988856</td><td>0.958027</td><td>0.913358</td><td>0.007332</td><td>0.003147</td><td>0.834222</td><td>0.998724</td><td>SEKER</td><td></td></tr><tr><td>1</td><td>28734</td><td>638.018</td><td>200.524796</td><td>182.734419</td><td>1.097356</td><td>0.411785</td><td>29172</td><td>191.272751</td><td>0.783968</td><td>0.984986</td><td>0.887034</td><td>0.933861</td><td>0.006979</td><td>0.003564</td><td>0.909851</td><td>0.998430</td><td>SEKER</td><td></td></tr><tr><td>2</td><td>29380</td><td>624.110</td><td>212.826130</td><td>175.931143</td><td>1.209713</td><td>0.560727</td><td>29680</td><td>193.410904</td><td>0.778113</td><td>0.989559</td><td>0.947849</td><td>0.908774</td><td>0.007244</td><td>0.003048</td><td>0.823871</td><td>0.999068</td><td>SEKER</td><td></td></tr><tr><td>3</td><td>30008</td><td>645.884</td><td>210.557999</td><td>182.510516</td><td>1.153608</td><td>0.498616</td><td>30724</td><td>195.467082</td><td>0.782681</td><td>0.976496</td><td>0.903936</td><td>0.928329</td><td>0.007017</td><td>0.003215</td><td>0.881754</td><td>0.994199</td><td>SEKER</td><td></td></tr><tr><td>4</td><td>30140</td><td>620.134</td><td>201.847882</td><td>190.279279</td><td>1.060798</td><td>0.333680</td><td>30417</td><td>195.896503</td><td>0.773098</td><td>0.950893</td><td>0.984877</td><td>0.970516</td><td>0.006697</td><td>0.003865</td><td>0.941900</td><td>0.999166</td><td>SEKER</td><td></td></tr><tr><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td></tr><tr><td>13606</td><td>42097</td><td>759.686</td><td>288.721812</td><td>185.944705</td><td>1.552728</td><td>0.768002</td><td>42508</td><td>231.515799</td><td>0.714574</td><td>0.990331</td><td>0.916603</td><td>0.801865</td><td>0.006858</td><td>0.001749</td><td>0.642988</td><td>0.998385</td><td>DERMAISON</td><td></td></tr><tr><td>13607</td><td>42101</td><td>757.499</td><td>281.576362</td><td>190.713136</td><td>1.476459</td><td>0.735702</td><td>42484</td><td>231.526798</td><td>0.799943</td><td>0.990752</td><td>0.922015</td><td>0.822252</td><td>0.006688</td><td>0.001886</td><td>0.676099</td><td>0.998219</td><td>DERMAISON</td><td></td></tr><tr><td>13608</td><td>42139</td><td>759.321</td><td>281.530928</td><td>191.187979</td><td>1.472582</td><td>0.734005</td><td>42589</td><td>231.631261</td><td>0.729932</td><td>0.989899</td><td>0.918424</td><td>0.822730</td><td>0.006681</td><td>0.001888</td><td>0.678884</td><td>0.996767</td><td>DERMAISON</td><td></td></tr><tr><td>13609</td><td>42147</td><td>763.779</td><td>283.382636</td><td>190.275731</td><td>1.489326</td><td>0.741035</td><td>42667</td><td>231.653247</td><td>0.705389</td><td>0.987813</td><td>0.907906</td><td>0.817457</td><td>0.006724</td><td>0.001852</td><td>0.688237</td><td>0.995222</td><td>DERMAISON</td><td></td></tr><tr><td>13610</td><td>42159</td><td>772.237</td><td>295.142741</td><td>182.204716</td><td>1.619841</td><td>0.786693</td><td>42600</td><td>231.686223</td><td>0.788962</td><td>0.989648</td><td>0.888380</td><td>0.784997</td><td>0.007001</td><td>0.001640</td><td>0.616221</td><td>0.998180</td><td>DERMAISON</td><td></td></tr></tbody></table> <div>13611 rows × 18 columns</div>		Area	Perimeter	MajorAxisLength	MinorAxisLength	AspectRatio	Eccentricity	ConvexArea	EquivalentDiameter	Extent	Solidity	roundness	Compactness	Shapefactor1	Shapefactor2	Shapefactor3	Shapefactor4	Class	Class	0	28395	610.291	208.178117	173.888747	1.197191	0.548812	28715	190.141097	0.763923	0.988856	0.958027	0.913358	0.007332	0.003147	0.834222	0.998724	SEKER		1	28734	638.018	200.524796	182.734419	1.097356	0.411785	29172	191.272751	0.783968	0.984986	0.887034	0.933861	0.006979	0.003564	0.909851	0.998430	SEKER		2	29380	624.110	212.826130	175.931143	1.209713	0.560727	29680	193.410904	0.778113	0.989559	0.947849	0.908774	0.007244	0.003048	0.823871	0.999068	SEKER		3	30008	645.884	210.557999	182.510516	1.153608	0.498616	30724	195.467082	0.782681	0.976496	0.903936	0.928329	0.007017	0.003215	0.881754	0.994199	SEKER		4	30140	620.134	201.847882	190.279279	1.060798	0.333680	30417	195.896503	0.773098	0.950893	0.984877	0.970516	0.006697	0.003865	0.941900	0.999166	SEKER		13606	42097	759.686	288.721812	185.944705	1.552728	0.768002	42508	231.515799	0.714574	0.990331	0.916603	0.801865	0.006858	0.001749	0.642988	0.998385	DERMAISON		13607	42101	757.499	281.576362	190.713136	1.476459	0.735702	42484	231.526798	0.799943	0.990752	0.922015	0.822252	0.006688	0.001886	0.676099	0.998219	DERMAISON		13608	42139	759.321	281.530928	191.187979	1.472582	0.734005	42589	231.631261	0.729932	0.989899	0.918424	0.822730	0.006681	0.001888	0.678884	0.996767	DERMAISON		13609	42147	763.779	283.382636	190.275731	1.489326	0.741035	42667	231.653247	0.705389	0.987813	0.907906	0.817457	0.006724	0.001852	0.688237	0.995222	DERMAISON		13610	42159	772.237	295.142741	182.204716	1.619841	0.786693	42600	231.686223	0.788962	0.989648	0.888380	0.784997	0.007001	0.001640	0.616221	0.998180	DERMAISON	
		Area	Perimeter	MajorAxisLength	MinorAxisLength	AspectRatio	Eccentricity	ConvexArea	EquivalentDiameter	Extent	Solidity	roundness	Compactness	Shapefactor1	Shapefactor2	Shapefactor3	Shapefactor4	Class	Class																																																																																																																																																																																																																		
	0	28395	610.291	208.178117	173.888747	1.197191	0.548812	28715	190.141097	0.763923	0.988856	0.958027	0.913358	0.007332	0.003147	0.834222	0.998724	SEKER																																																																																																																																																																																																																			
	1	28734	638.018	200.524796	182.734419	1.097356	0.411785	29172	191.272751	0.783968	0.984986	0.887034	0.933861	0.006979	0.003564	0.909851	0.998430	SEKER																																																																																																																																																																																																																			
	2	29380	624.110	212.826130	175.931143	1.209713	0.560727	29680	193.410904	0.778113	0.989559	0.947849	0.908774	0.007244	0.003048	0.823871	0.999068	SEKER																																																																																																																																																																																																																			
	3	30008	645.884	210.557999	182.510516	1.153608	0.498616	30724	195.467082	0.782681	0.976496	0.903936	0.928329	0.007017	0.003215	0.881754	0.994199	SEKER																																																																																																																																																																																																																			
	4	30140	620.134	201.847882	190.279279	1.060798	0.333680	30417	195.896503	0.773098	0.950893	0.984877	0.970516	0.006697	0.003865	0.941900	0.999166	SEKER																																																																																																																																																																																																																			
																																																																																																																																																																																																																		
	13606	42097	759.686	288.721812	185.944705	1.552728	0.768002	42508	231.515799	0.714574	0.990331	0.916603	0.801865	0.006858	0.001749	0.642988	0.998385	DERMAISON																																																																																																																																																																																																																			
	13607	42101	757.499	281.576362	190.713136	1.476459	0.735702	42484	231.526798	0.799943	0.990752	0.922015	0.822252	0.006688	0.001886	0.676099	0.998219	DERMAISON																																																																																																																																																																																																																			
13608	42139	759.321	281.530928	191.187979	1.472582	0.734005	42589	231.631261	0.729932	0.989899	0.918424	0.822730	0.006681	0.001888	0.678884	0.996767	DERMAISON																																																																																																																																																																																																																				
13609	42147	763.779	283.382636	190.275731	1.489326	0.741035	42667	231.653247	0.705389	0.987813	0.907906	0.817457	0.006724	0.001852	0.688237	0.995222	DERMAISON																																																																																																																																																																																																																				
13610	42159	772.237	295.142741	182.204716	1.619841	0.786693	42600	231.686223	0.788962	0.989648	0.888380	0.784997	0.007001	0.001640	0.616221	0.998180	DERMAISON																																																																																																																																																																																																																				
Univariate Analysis																																																																																																																																																																																																																																					



Multivariate Analysis



Bivariate Analysis

-

Outliers and Anomalies

-

Data Preprocessing Code Screenshots

Loading Data

```
df=pd.read_csv('content/Bry.Bean.Datasets.csv')
print(df.head(10))
print(df.info())

# Convert 'Class' column to numerical representation if it contains categorical data
df['Class_Numeric'] = df['Class'].astype('category').cat.codes

# Calculate correlation matrix excluding non-numerical columns
corr_matrix = df.select_dtypes(include='number').corr()
```

	Area	Perimeter	MajorAxisLength	MinorAxisLength	AspectRatio	Eccentricity	ConvexArea	EquiDiameter	Extent	Solidity	Roundness	Compactness	Shapefactor1	Shapefactor2	Shapefactor3	Shapefactor4	Class	Class
0	28395	610291	258.178117	173.888747	1.497191	0.549812	28715	190.141097	0.743925	0.988856	0.658027	0.613558	0.007352	0.003147	0.834222	0.998724	SEVER	
1	28734	608318	255.521796	182.756619	1.387556	0.611785	29172	191.372751	0.783968	0.984096	0.887034	0.655061	0.006979	0.003564	0.900551	0.998480	SEVER	
2	29380	624110	272.829320	175.931142	1.559713	0.582727	29690	193.410904	0.731133	0.989359	0.947849	0.698774	0.007244	0.003348	0.823671	0.999080	SEVER	
3	30008	645884	218.537999	182.510516	1.153038	0.498616	30724	195.407002	0.782661	0.976094	0.903936	0.928529	0.007017	0.003215	0.841794	0.994199	SEVER	
4	30140	620134	251.847882	190.270279	1.300798	0.533680	30417	195.894503	0.773608	0.990993	0.864877	0.970316	0.006697	0.003065	0.941503	0.999166	SEVER	
...
13608	42097	790308	288.771812	180.944709	1.597708	0.708002	42038	231.811199	0.718314	0.990317	0.979603	0.907193	0.008908	0.007199	0.942089	0.999393	DEMAISON	
13607	42101	787495	281.574592	190.718136	1.476438	0.739702	42434	231.826798	0.799940	0.991792	0.922015	0.906668	0.009866	0.009866	0.970099	0.998219	DEMAISON	
13608	42130	790307	285.509928	191.181919	1.472982	0.736005	42989	231.801261	0.759602	0.989999	0.979424	0.922702	0.008801	0.007988	0.978684	0.998767	DEMAISON	
13609	42147	793779	283.380308	190.218731	1.489028	0.740305	42987	231.850247	0.759389	0.989103	0.967908	0.917487	0.008724	0.007982	0.988037	0.999222	DEMAISON	
13610	42150	772237	295.142741	182.204716	1.619541	0.786885	42500	231.686223	0.789962	0.989548	0.888380	0.784997	0.007001	0.006140	0.616221	0.998180	DEMAISON	

13611 rows x 18 columns

Managing Missing Data	<pre>df.isna().sum()</pre> <pre>Area 0 Perimeter 0 MajorAxisLength 0 MinorAxisLength 0 AspectRatio 0 Eccentricity 0 ConvexArea 0 EquivDiameter 0 Extent 0 Solidity 0 roundness 0 Compactness 0 ShapeFactor1 0 ShapeFactor2 0 ShapeFactor3 0 ShapeFactor4 0 Class 0 Class_Numeric 0 dtype: int64</pre>																																																																																																																																																																																																												
Data Transformation	<pre>df.describe(include="all")</pre> <table><tr><th></th><th>Area</th><th>Perimeter</th><th>MajorAxisLength</th><th>MinorAxisLength</th><th>AspectRatio</th><th>Eccentricity</th><th>ConvexArea</th><th>EquivDiameter</th><th>Extent</th><th>Solidity</th><th>roundness</th><th>Compactness</th><th>Shapefactor1</th><th>Shapefactor2</th><th>Shapefactor3</th><th>Shapefactor4</th></tr><tr><td>count</td><td>12934.000000</td><td>12934.000000</td><td>12934.000000</td><td>12934.000000</td><td>12934.000000</td><td>12934.000000</td><td>12934.000000</td><td>12934.000000</td><td>12934.000000</td><td>12934.000000</td><td>12934.000000</td><td>12934.000000</td><td>12934.000000</td><td>12934.000000</td><td>12934.000000</td><td>12934.000000</td></tr><tr><td>unique</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td></tr><tr><td>top</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td></tr><tr><td>freq</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td><td>NaN</td></tr><tr><td>mean</td><td>47675.713666</td><td>821.806271</td><td>357.499513</td><td>194.523583</td><td>1.381764</td><td>0.749470</td><td>48324.716329</td><td>243.208630</td><td>0.748594</td><td>0.987191</td><td>0.876324</td><td>0.800367</td><td>0.006711</td><td>0.001799</td><td>0.644849</td><td>0.1</td></tr><tr><td>std</td><td>15711.661197</td><td>153.703140</td><td>64.933076</td><td>28.212878</td><td>0.251235</td><td>0.093057</td><td>16016.910449</td><td>39.403635</td><td>0.040215</td><td>0.050405</td><td>0.060378</td><td>0.062786</td><td>0.000940</td><td>0.000578</td><td>0.100794</td><td>0.1</td></tr><tr><td>min</td><td>35402.000000</td><td>524.736000</td><td>183.601165</td><td>122.812653</td><td>1.024868</td><td>0.218951</td><td>20584.000000</td><td>161.243764</td><td>0.585315</td><td>0.819248</td><td>0.489618</td><td>0.440377</td><td>0.004241</td><td>0.000791</td><td>0.410359</td><td>0.1</td></tr><tr><td>25%</td><td>35902.250000</td><td>699.271750</td><td>231.560461</td><td>174.717207</td><td>1.426325</td><td>0.713064</td><td>36317.750000</td><td>213.854033</td><td>0.717352</td><td>0.985765</td><td>0.830929</td><td>0.762441</td><td>0.006093</td><td>0.001217</td><td>0.581316</td><td>0.1</td></tr><tr><td>50%</td><td>43692.500000</td><td>784.225000</td><td>291.857544</td><td>191.049028</td><td>1.547315</td><td>0.763100</td><td>44201.500000</td><td>235.862288</td><td>0.758600</td><td>0.988308</td><td>0.885686</td><td>0.802403</td><td>0.006683</td><td>0.001741</td><td>0.643851</td><td>0.1</td></tr><tr><td>75%</td><td>57351.750000</td><td>946.790750</td><td>367.346394</td><td>209.798897</td><td>1.707675</td><td>0.810006</td><td>58238.500000</td><td>270.226786</td><td>0.783974</td><td>0.990006</td><td>0.918438</td><td>0.835795</td><td>0.007320</td><td>0.002200</td><td>0.698553</td><td>0.1</td></tr><tr><td>max</td><td>90231.000000</td><td>1273.205000</td><td>479.134377</td><td>301.705356</td><td>2.420206</td><td>0.911423</td><td>91438.000000</td><td>338.947898</td><td>0.866195</td><td>0.994378</td><td>0.990685</td><td>0.987303</td><td>0.010451</td><td>0.003665</td><td>0.974767</td><td>0.1</td></tr></table>		Area	Perimeter	MajorAxisLength	MinorAxisLength	AspectRatio	Eccentricity	ConvexArea	EquivDiameter	Extent	Solidity	roundness	Compactness	Shapefactor1	Shapefactor2	Shapefactor3	Shapefactor4	count	12934.000000	12934.000000	12934.000000	12934.000000	12934.000000	12934.000000	12934.000000	12934.000000	12934.000000	12934.000000	12934.000000	12934.000000	12934.000000	12934.000000	12934.000000	12934.000000	unique	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	top	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	freq	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	mean	47675.713666	821.806271	357.499513	194.523583	1.381764	0.749470	48324.716329	243.208630	0.748594	0.987191	0.876324	0.800367	0.006711	0.001799	0.644849	0.1	std	15711.661197	153.703140	64.933076	28.212878	0.251235	0.093057	16016.910449	39.403635	0.040215	0.050405	0.060378	0.062786	0.000940	0.000578	0.100794	0.1	min	35402.000000	524.736000	183.601165	122.812653	1.024868	0.218951	20584.000000	161.243764	0.585315	0.819248	0.489618	0.440377	0.004241	0.000791	0.410359	0.1	25%	35902.250000	699.271750	231.560461	174.717207	1.426325	0.713064	36317.750000	213.854033	0.717352	0.985765	0.830929	0.762441	0.006093	0.001217	0.581316	0.1	50%	43692.500000	784.225000	291.857544	191.049028	1.547315	0.763100	44201.500000	235.862288	0.758600	0.988308	0.885686	0.802403	0.006683	0.001741	0.643851	0.1	75%	57351.750000	946.790750	367.346394	209.798897	1.707675	0.810006	58238.500000	270.226786	0.783974	0.990006	0.918438	0.835795	0.007320	0.002200	0.698553	0.1	max	90231.000000	1273.205000	479.134377	301.705356	2.420206	0.911423	91438.000000	338.947898	0.866195	0.994378	0.990685	0.987303	0.010451	0.003665	0.974767	0.1
	Area	Perimeter	MajorAxisLength	MinorAxisLength	AspectRatio	Eccentricity	ConvexArea	EquivDiameter	Extent	Solidity	roundness	Compactness	Shapefactor1	Shapefactor2	Shapefactor3	Shapefactor4																																																																																																																																																																																													
count	12934.000000	12934.000000	12934.000000	12934.000000	12934.000000	12934.000000	12934.000000	12934.000000	12934.000000	12934.000000	12934.000000	12934.000000	12934.000000	12934.000000	12934.000000	12934.000000																																																																																																																																																																																													
unique	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN																																																																																																																																																																																													
top	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN																																																																																																																																																																																													
freq	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN																																																																																																																																																																																													
mean	47675.713666	821.806271	357.499513	194.523583	1.381764	0.749470	48324.716329	243.208630	0.748594	0.987191	0.876324	0.800367	0.006711	0.001799	0.644849	0.1																																																																																																																																																																																													
std	15711.661197	153.703140	64.933076	28.212878	0.251235	0.093057	16016.910449	39.403635	0.040215	0.050405	0.060378	0.062786	0.000940	0.000578	0.100794	0.1																																																																																																																																																																																													
min	35402.000000	524.736000	183.601165	122.812653	1.024868	0.218951	20584.000000	161.243764	0.585315	0.819248	0.489618	0.440377	0.004241	0.000791	0.410359	0.1																																																																																																																																																																																													
25%	35902.250000	699.271750	231.560461	174.717207	1.426325	0.713064	36317.750000	213.854033	0.717352	0.985765	0.830929	0.762441	0.006093	0.001217	0.581316	0.1																																																																																																																																																																																													
50%	43692.500000	784.225000	291.857544	191.049028	1.547315	0.763100	44201.500000	235.862288	0.758600	0.988308	0.885686	0.802403	0.006683	0.001741	0.643851	0.1																																																																																																																																																																																													
75%	57351.750000	946.790750	367.346394	209.798897	1.707675	0.810006	58238.500000	270.226786	0.783974	0.990006	0.918438	0.835795	0.007320	0.002200	0.698553	0.1																																																																																																																																																																																													
max	90231.000000	1273.205000	479.134377	301.705356	2.420206	0.911423	91438.000000	338.947898	0.866195	0.994378	0.990685	0.987303	0.010451	0.003665	0.974767	0.1																																																																																																																																																																																													
Feature Engineering	Attached the codes in final submission.																																																																																																																																																																																																												
Save Processed Data	-																																																																																																																																																																																																												