```
1 D:\Project\Amazon\.venv\Scripts\python.exe D:\Project
   \Amazon\main.pv
 2 *******Amazon Delivery Time Prediction
   *****
 3 ******Step 1:- Load DataSet********
 4
          Order_ID Agent_Age ... Delivery_Time
   Category
 5 0 ialx566343618
                            37
                                              120
   Clothing
 6 1 akqg208421122
                            34
                                              165
   Electronics
 7 2 njpu434582536
                            23
                                              130
          Sports
 8 3 rjto796129700
                            38
                                              105
   Cosmetics
 9 4 zguw716275638
                            32
                                              150
            Toys
10
11 [5 rows x 16 columns]
12 <class 'pandas.core.frame.DataFrame'>
13 RangeIndex: 43739 entries, 0 to 43738
14 Data columns (total 16 columns):
15
   #
                         Non-Null Count
        Column
                                        Dtype
16 ---
        _ _ _ _ _
                                         ----
17
   0
        Order_ID
                        43739 non-null
                                        object
18
       Agent_Age
                        43739 non-null
                                        int64
   1
19
   2
       Agent_Rating
                        43685 non-null
                                        float64
20
    3
       Store_Latitude
                        43739 non-null
                                        float64
21
    4
       Store_Longitude
                        43739 non-null
                                        float64
22
    5
       Drop_Latitude
                        43739 non-null
                                        float64
23
    6
        Drop_Longitude
                        43739 non-null
                                        float64
24
    7
       Order_Date
                        43739 non-null
                                        object
25
                        43739 non-null
    8
        Order_Time
                                        object
26
   9
       Pickup_Time
                        43739 non-null
                                        object
27
   10
       Weather
                        43648 non-null
                                        object
28
   11
       Traffic
                        43739 non-null
                                        object
29
   12
       Vehicle
                        43739 non-null
                                        object
30
   13
       Area
                        43739 non-null
                                        object
31
   14
       Delivery_Time
                        43739 non-null
                                        int64
32
    15
                        43739 non-null
        Category
                                        object
33 dtypes: float64(5), int64(2), object(9)
```

File - r	nain				
34	memory usage: 5.3+ MB				
35	None				
36	Agent_Age A	\gent_Rating	• • •		
	Drop_Longitude Delive	ery_Time			
37	count 43739.000000 4	3685.000000	• • •	43739.	
	000000 43739.000000				
38	mean 29.567137	4.633780		70.	
	821842 124.905645				
39	std 5.815155		• • •	21.	
	153148 51.915451				
40	min 15.000000	1.000000		0.	
	010000 10.000000				
41	25% 25.000000	4.500000	• • •	73.	
	280000 90.000000				
42	50% 30.000000	4.700000	• • •	76.	
	002574 125.000000				
43	75% 35.000000	4.900000		78.	
	104095 160.000000				
44	max 50.000000	6.000000		88.	
	563452 270.000000				
45					
46	[8 rows x 7 columns]				
47	Original Columns ['Oro	ler_ID', 'Age	nt_Age	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	
	Agent_Rating', 'Store_Latitude', 'Store_Longitude', '				
	<pre>Drop_Latitude', 'Drop_Longitude', 'Order_Date', '</pre>				
	Order_Time', 'Pickup_Time', 'Weather', 'Traffic', '				
	Vehicle', 'Area', 'Del	.ivery_Time',	'Cate	gory']	
48	********Step 2A :- H	landling miss	ing va	lues******	
49	Missi	g Values %M	issing	Percentage	
50	Order_ID	0		0.000000	
51	Agent_Age	0		0.000000	
52	Agent_Rating	54		0.123460	
53	Store_Latitude	0		0.000000	
54	Store_Longitude	0		0.000000	
55	Drop_Latitude	0		0.00000	
56	Drop_Longitude	0		0.00000	
57	Order_Date	0		0.00000	
58	Order_Time	0		0.00000	
59	Pickup_Time	0		0.00000	
60	Weather	91		0.208052	
61	Traffic	0		0.00000	

```
62 Vehicle
                                0
                                               0.000000
                                0
63 Area
                                               0.000000
64 Delivery_Time
                                0
                                               0.000000
65 Category
                                0
                                               0.000000
66 df2 = Case 1 Coulumn is 100% missing
67
           Order_ID Agent_Age
                                . . .
                                     Delivery_Time
   Category
68 0 ialx566343618
                            37
                                                120
   Clothing
69 1 akqq208421122
                            34
                                                165
   Electronics
70 2 njpu434582536
                            23
                                                130
          Sports
71 3 rjto796129700
                            38
                                                105
   Cosmetics
72 4 zguw716275638
                            32
                                                150
            Toys
73
74 [5 rows x 16 columns]
75 Columns After Removing Missing Values= ['Order_ID
   ', 'Agent_Age', 'Agent_Rating', 'Store_Latitude', '
   Store_Longitude', 'Drop_Latitude', 'Drop_Longitude
   ', 'Order_Date', 'Order_Time', 'Pickup_Time', '
  Weather', 'Traffic', 'Vehicle', 'Area', '
   Delivery_Time', 'Category']
76 Original Columns ['Order_ID', 'Agent_Age', '
   Agent_Rating', 'Store_Latitude', 'Store_Longitude
   ', 'Drop_Latitude', 'Drop_Longitude', 'Order_Date
      'Order_Time', 'Pickup_Time', 'Weather', 'Traffic
   ', 'Vehicle', 'Area', 'Delivery_Time', 'Category']
77 After comparing it is conclude that dataset is
   having zero column having 100% missing value
78 Case 2 Column Is mostly missing >50%
79 ********Step 2B:- Fill Missing Values*******
80 Agent_rating(Numeric Column) == using median(safer
   than mean if outliers exist
81 Weather(Categorical Column) == using mode
82 Order_ID
                      0
83 Agent_Age
                      0
84 Agent_Rating
                      0
85 Store_Latitude
                      0
```

```
86 Store_Longitude
                        0
 87 Drop_Latitude
                        0
 88 Drop_Longitude
                        0
 89 Order_Date
                        0
 90 Order_Time
                        0
 91 Pickup_Time
                        0
 92 Weather
                        0
 93 Traffic
                        0
 94 Vehicle
                        0
 95 Area
                        0
 96 Delivery_Time
                        0
 97 Category
                        0
 98 dtype: int64
 99 Missing values handled successfully!
100 Remaining Nulls After Filling:
101 Order_ID
                        0
102 Agent_Age
                        0
103 Agent_Rating
                        0
104 Store_Latitude
                        0
105 Store_Longitude
                        0
106 Drop_Latitude
                        0
107 Drop_Longitude
                        0
108 Order Date
                        0
109 Order_Time
                        0
110 Pickup_Time
                        0
111 Weather
                        0
112 Traffic
                        0
113 Vehicle
                        0
114 Area
                        0
115 Delivery_Time
                        0
116 Category
                        0
117 dtype: int64
118 ***********Step 3 Handling Duplicates
    *********
119 Total Duplicates:
120 Empty DataFrame
121 Columns: [Order_ID, Agent_Age, Agent_Rating,
    Store_Latitude, Store_Longitude, Drop_Latitude,
    Drop_Longitude, Order_Date, Order_Time, Pickup_Time
    , Weather, Traffic, Vehicle, Area, Delivery_Time,
    Category]
```

File - ma	lin					
122	<pre>Index: []</pre>					
123	NO duplicates found.					
124	********Step 4 Feature Engineering******					
125	Euclidean distance is en	ough unless	exact km/miles			
	are needed.					
126	Store_Latitude Store	_Longitude	Drop_Latitude			
	Distance	_	, –			
127	0 22.745049	75.892471	22.765049 0			
	.028284					
	1 12.913041	77.683237	13.043041 0			
	.183848					
	2 12.914264	77.678400	12.924264 0			
	.014142	77.070100	12.72 1201			
	3 11.003669	76.976494	11.053669 0			
	.070711	70.770474	11.000007			
	4 12.972793	80 2/QQ82	13.012793 0			
131	.056569	00.247702	13.012773 0			
132	=========Histogram	of Fueledes	un distance			
132	======================================	or Loctedes	III utstance			
177	Step 4B: Haversine gives	a poalistic	foatung instead			
133	of a flat approximation					
	accuracy.	. IIII Call III	ipi ove modec 3			
13/	Distance_km					
	0 3.025149					
l .	1 20.183530					
	2 1.552758					
	3 7.790401					
	4 6.210138					
		om of Howens	ina diatanasa (
140	O =================Histogram of Haversine distances (
1/1	in km)====================================		Faaiaaaiaa			
141	*******Step 5: Time-Ba	ased Feature	Engineering			
4 (0	****					
142		day Is_Week	cend Order_Hour			
	Pickup_Hour Rush_Hour					
143	0 2022-03-19	5	1 11.0			
	11 0					
144	1 2022-03-25	4	0 19.0			
	19 1					
145	2 2022-03-19	5	1 8.0			
	8 1					
146	3 2022-04-05	1	0 18.0			

```
146
               18
                           1
147 4 2022-03-26
                              5
                                                   13.0
                                          1
               13
                           0
148 ********Step 6: Pickup Delay Feature******
149 Pickup Delay feature measures how long it takes for
    the agent to actually pick up the parcel after the
    order is placed.
150
           Order_DateTime
                              Pickup_DateTime
   Pickup_Delay_Minutes
151 0 2022-03-19 11:30:00 2022-03-19 11:45:00
                      15.0
152 1 2022-03-25 19:45:00 2022-03-25 19:50:00
                       5.0
153 2 2022-03-19 08:30:00 2022-03-19 08:45:00
                      15.0
154 3 2022-04-05 18:00:00 2022-04-05 18:10:00
                      10.0
155 4 2022-03-26 13:30:00 2022-03-26 13:45:00
                      15.0
156 **********Step 7: one hot encoding*******
157 Train Score: 0.9729247483785157
158 Test Score: 0.8091544277212389
159 Step 8 :- EDA
160 Boxplot to check outliers
161 Scatter plot vs delivery time
162 ======Distribution of Delivery_Time (target
    variable)=======
163 =======Numeric Features: Agent_Age, Agent_Rating
    , Distance_km, Pickup_Delay_Minutes
164 Boxplots to detect outliers
165 ======Correlation Heatmap=======
166 Categorical Features Visualization
167 D:\Project\Amazon\main.py:211: FutureWarning:
168
169 Passing `palette` without assigning `hue` is
    deprecated and will be removed in v0.14.0. Assign
    the `x` variable to `hue` and set `legend=False` for
     the same effect.
170
      sns.countplot(data=df, x=col, order=df[col].
171
   value_counts().index, palette='Set2')
```

```
172 D:\Project\Amazon\main.py:211: FutureWarning:
173
174 Passing `palette` without assigning `hue` is
    deprecated and will be removed in v0.14.0. Assign
    the `x` variable to `hue` and set `legend=False` for
     the same effect.
175
176
      sns.countplot(data=df, x=col, order=df[col].
    value_counts().index, palette='Set2')
177 D:\Project\Amazon\main.py:211: FutureWarning:
178
179 Passing `palette` without assigning `hue` is
    deprecated and will be removed in v0.14.0. Assign
    the `x` variable to `hue` and set `legend=False` for
     the same effect.
180
      sns.countplot(data=df, x=col, order=df[col].
181
    value_counts().index, palette='Set2')
182 D:\Project\Amazon\main.py:211: FutureWarning:
183
184 Passing `palette` without assigning `hue` is
    deprecated and will be removed in v0.14.0. Assign
    the `x` variable to `hue` and set `legend=False` for
     the same effect.
185
186
      sns.countplot(data=df, x=col, order=df[col].
    value_counts().index, palette='Set2')
187 D:\Project\Amazon\main.py:211: FutureWarning:
188
189 Passing `palette` without assigning `hue` is
    deprecated and will be removed in v0.14.0. Assign
    the `x` variable to `hue` and set `legend=False` for
     the same effect.
190
191
      sns.countplot(data=df, x=col, order=df[col].
    value_counts().index, palette='Set2')
192 Feature Relationships
193
194 Process finished with exit code 0
195
```