



# ANALYSING THE RETAIL CUSTOMER RETENTION POTENTIAL LEVEL AS HIGH/MEDIUM/LOW

---

PRINCIPLES OF BIG DATA

# TEAM MEMBERS

---

- AVINASH CHOWDARY
- HARISH RAVILLA
- SAI PRANEETH
- SAI KIRAN
- RAVI CHANDRA THOTA
- VARUNDEV NUTALAPATI

# RENTENTIAL POTENTIAL

- Retail company draws the attention of the customers through captivating continual conversations for not facing customer churn or dropouts in the halfway.
- Retailers create the metric to find out the services that they need without losing them through analysis the historical data and once they are outdated from the plan, we have to try regain them back.
- We have to focus on medium level customers by satisfying their needs because they are accountable, proactive, reactive in terms of demanding.
- If they are satisfied, can renewal their plans and the continuity with the company So let's deep dive into the patterns and emphasizing concept further

# DATA SET

- The data set contains three types of details which are customers, advertise email and transactions history.
- The dataset is in the range of 2009 to 2018.

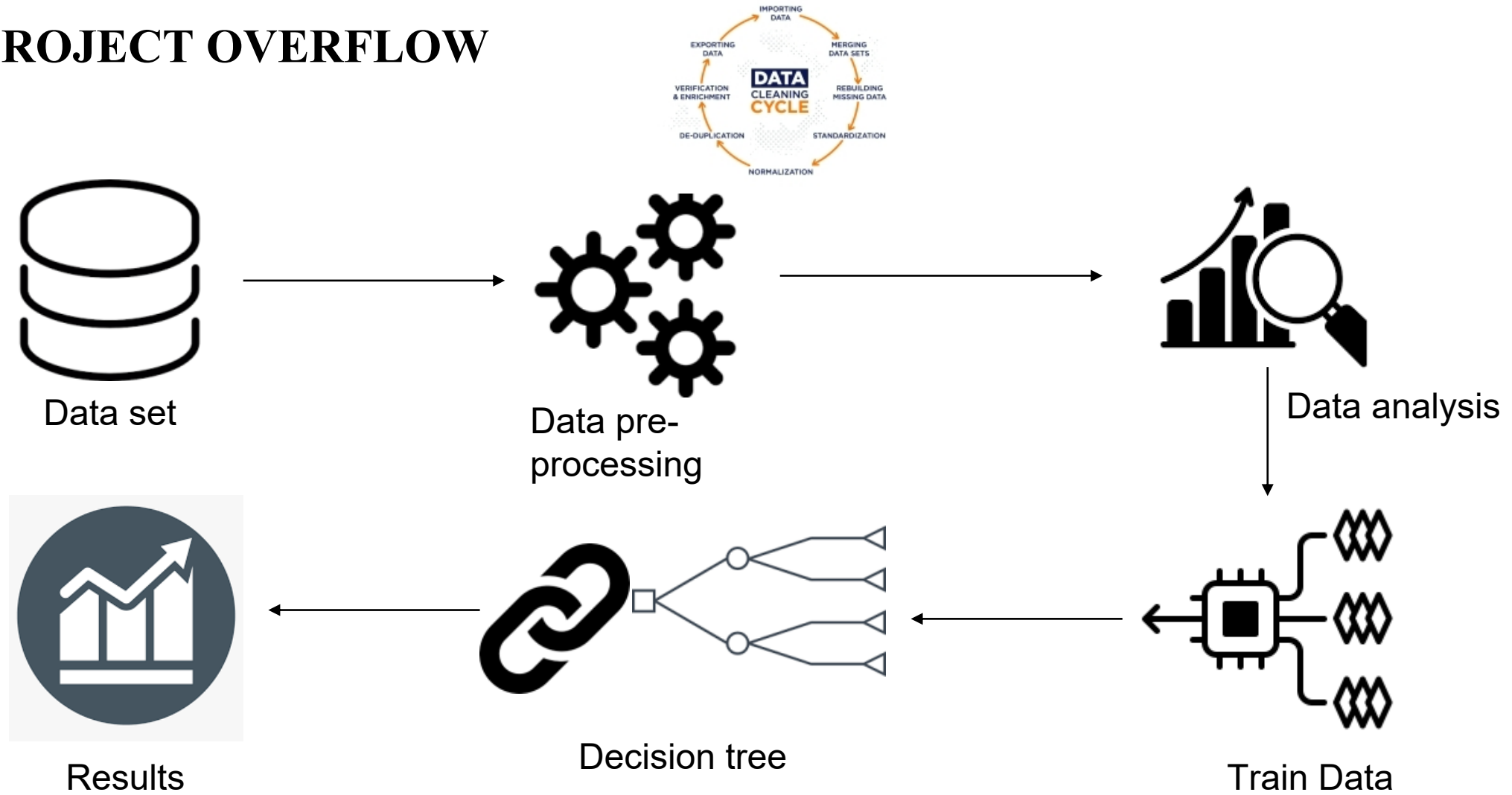
## Attributes

- The location details of the customer
  - The city from which the order is placed.
- Previous online communication details
  - the day they get the order
- Preferred delivery type and day
  - repeated order
- Automatic order subscriptions
  - total number of order
- Total Orders.
  - the frequency at which the person is ordering
- Order Frequency
  - Average number of days between his two orders
- Order gap
  - number of days from which he the company
- The number of days with business registered with
  -
- Business Profits
  - the customers interaction with the company
- Email responses
  - The signup date.
- Date of registration

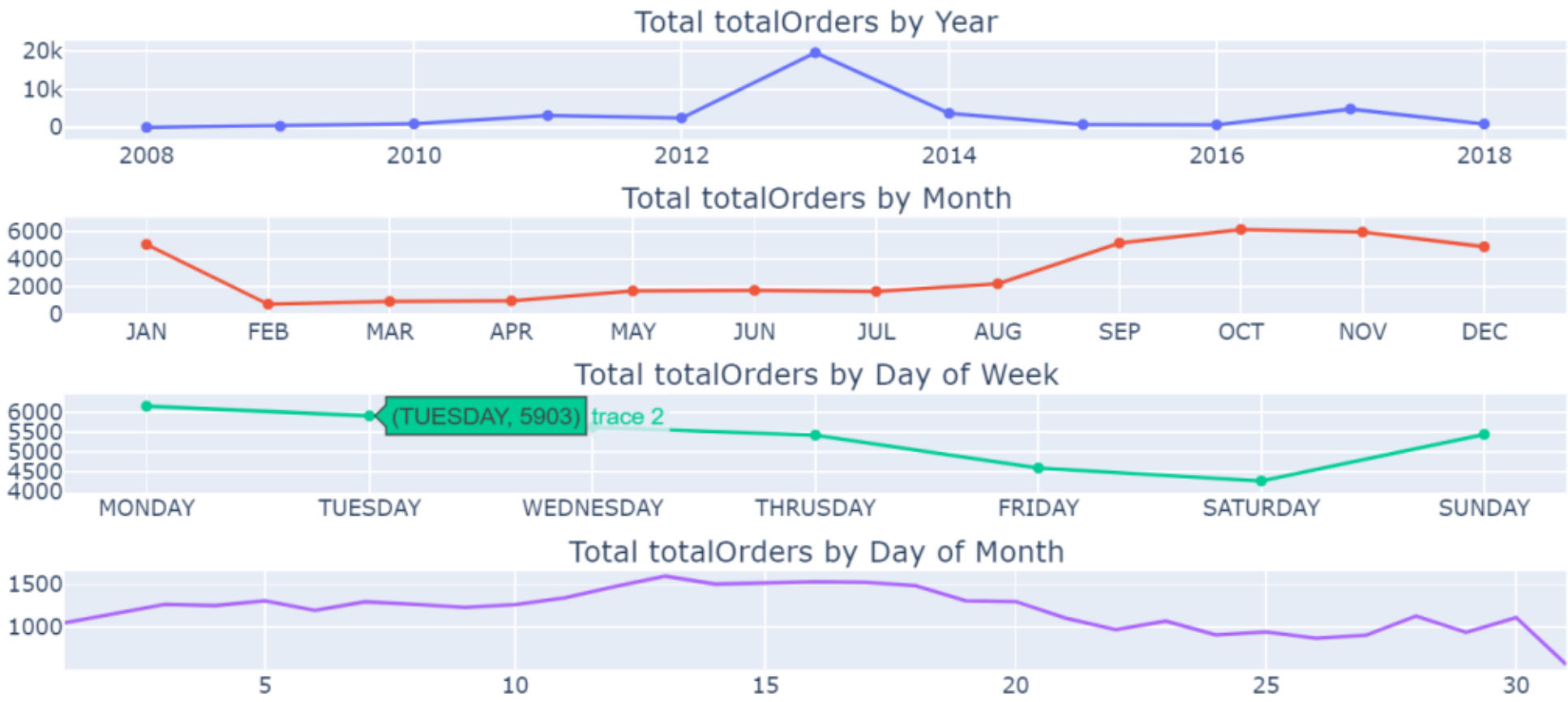
# DESCRIPTION OF DATA SET

A	B	C	D	E	F	G	O	P	Q	R	S	T	V	W	X
CustomerID	City	DateOfRegistration	OnlineCor	Automatic	Doorstep	Preferred	WelcomeEmailD	RecentInteract	totalOrder	OrderFreq	firstOrder	lastOrder	numberOf	makingBusi	ordergap
C16368	CITY2	9/28/2012	0	0	0	Monday	9/28/2012	7/25/2013	15	1	8/11/2013	8/11/2013	300	0	1624
C20345	CITY2	12/19/2010	1	1	1	Friday	12/19/2010	12/19/2013	15614	186	4/1/2011	1/19/2014	1096	1024	1463
C17683	CITY2	10/3/2010	0	0	0	Wednesday	10/3/2010	10/3/2010	443	13	12/1/2010	7/6/2011	0	217	2391
C18647	CITY4	11/27/2010	0	0	0	Monday	11/27/2010	12/17/2012	748	7	11/29/2010	1/28/2013	751	791	1819
C36692	CITY2	12/4/2009	1	1	0	Friday	12/4/2009	1/16/2014	2350	34	9/21/2011	1/16/2014	1504	848	1466
C36692	CITY2	12/4/2009	1	1	0	Friday	12/4/2009	1/16/2014	2350	34	9/21/2011	1/16/2014	1504	848	1466
C13831	CITY2	12/29/2010	1	0	0	Monday	12/29/2010	12/5/2013	9013	137	2/20/2011	12/10/2013	1072	1024	1503
C40881	CITY2	9/22/2010	0	0	1	Friday	9/22/2010	1/14/2014	8674	100	11/18/2010	1/15/2014	1210	1154	1467
C40881	CITY2	9/22/2010	0	0	1	Friday	9/22/2010	1/14/2014	8674	100	11/18/2010	1/15/2014	1210	1154	1467
C26246	CITY2	11/16/2009	0	0	0	Wednesday	11/16/2009	1/5/2014	7478	86	5/9/2011	1/5/2014	1511	972	1477
C26246	CITY2	11/16/2009	0	0	0	Wednesday	11/16/2009	1/5/2014	7478	86	5/9/2011	1/5/2014	1511	972	1477
C36136	CITY2	7/25/2009	1	1	1	Monday	7/25/2009	1/10/2014	8783	69	11/15/2010	1/19/2014	1630	1161	1463
C36136	CITY2	7/25/2009	1	1	1	Monday	7/25/2009	1/10/2014	8783	69	11/15/2010	1/19/2014	1630	1161	1463
C34046	CITY2	3/14/2010	0	0	0	Tuesday	3/14/2010	7/19/2013	8852	75	3/1/2011	10/6/2013	1223	950	1568
C34046	CITY2	3/14/2010	0	0	0	Tuesday	3/14/2010	7/19/2013	8852	75	3/1/2011	10/6/2013	1223	950	1568
C23215	CITY2	4/11/2011	1	1	0	Tuesday	4/11/2011	12/26/2013	7739	130	4/14/2011	1/13/2014	990	1005	1469
C24973	CITY4	3/8/2011	1	1	0	Wednesday	3/8/2011	1/5/2014	2256	45	3/27/2011	1/7/2014	1034	1017	1475
C12657	CITY2	8/30/2010	0	0	0	Monday	8/30/2010	10/12/2013	2209	22	3/12/2012	10/15/2013	1139	582	1559

# PROJECT OVERFLOW

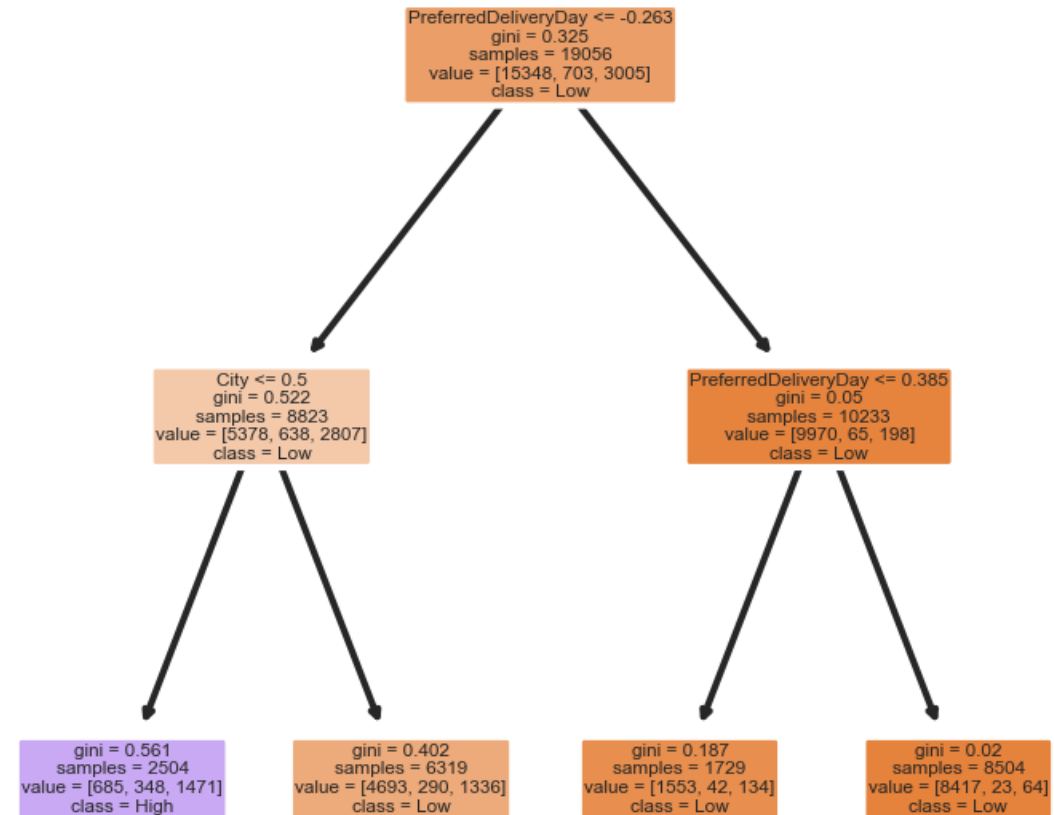


# DATA VISUALIZATION



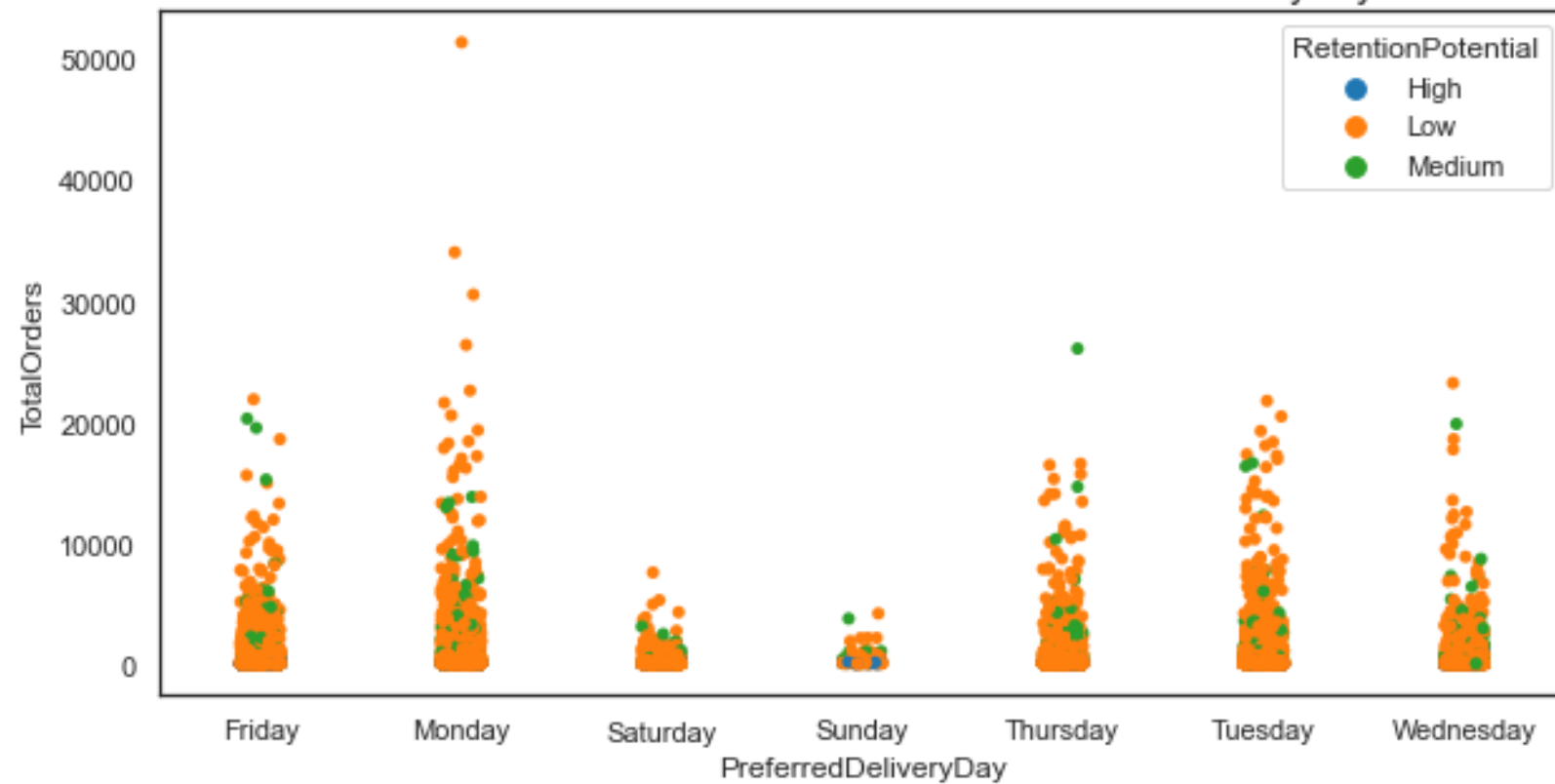
# DECISION TREE

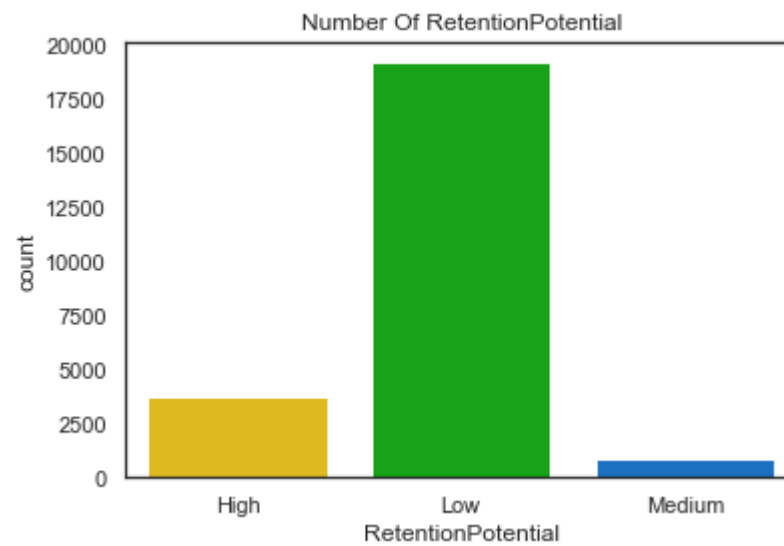
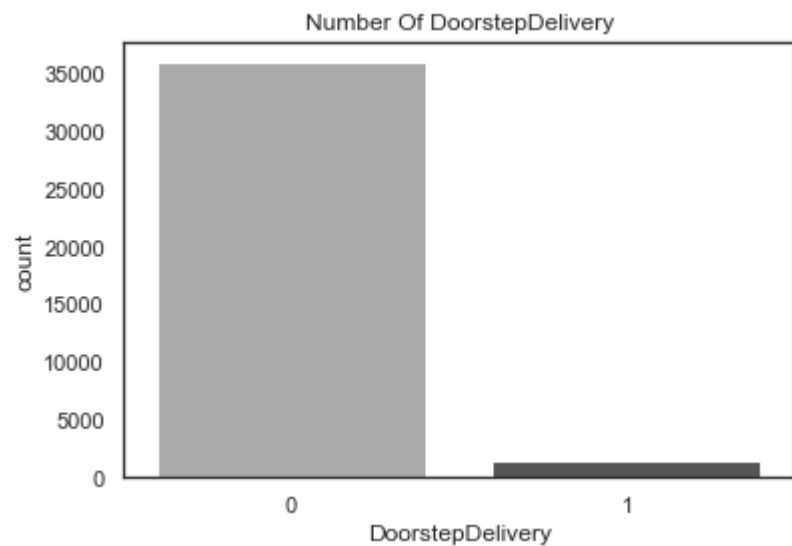
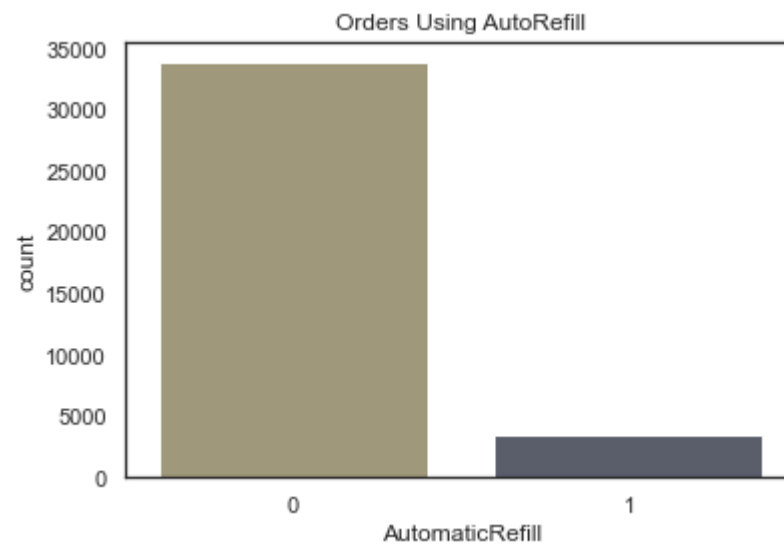
- Used normalization as a scaling technique in which values are shifted and rescaled so that they end up ranging between -1 and 1.
- Gini Index: It calculates the amount of probability of a specific feature that is classified incorrectly when selected randomly.



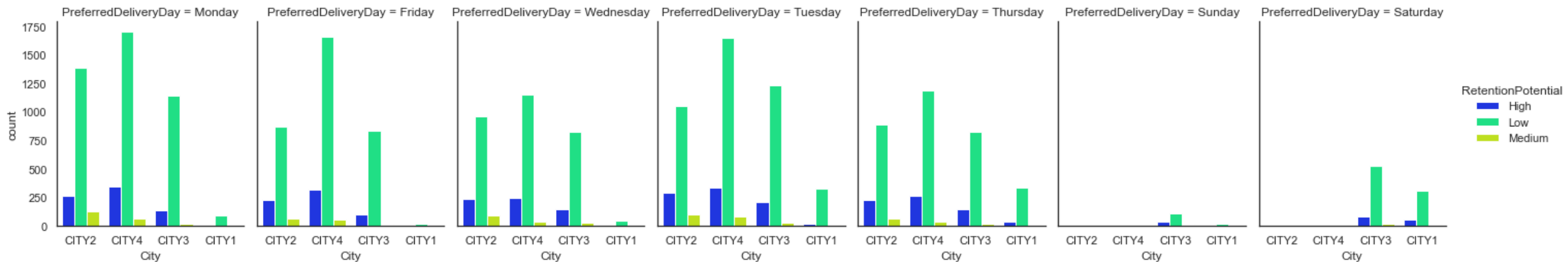


Distribution of TotalOrders variable with PreferredDeliveryDay





Comparison Between the City , PreferredDeliveryDay , Customer RetentionPotential



## CONCLUSION

- The customer's classification as LOW, MEDIUM, HIGH Retention Potential applies to all industries.
- To meet the profitability expectations without facing the condition of attrition,
  - Business should concentrate more on medium retention potential customers in which vigorous efforts will lead them to the high potentiality category.
  - Industries should fulfil the needs of customers having high potential, without any difficulties.
  - Another focus has to be on the people who tend to have low retention potential by improving the approaching analysis with captivating & fascinating attempts from industries to pull customers towards their services.