

EVENT COUPON BUNDLE



Predicting Coupon Acceptance on E-Commerce Platforms

Goal :

Finding How Many People Accept Coupons that will help in growing business in future

CONTENTS

- 01 **Introduction**
Why this project is and how machine learning can help in achieving our goal.
- 02 **Executive summary**
Summary of the Project that is all explanation in one page.
- 03 **ML model detail**
This section we will know about ml model and its result.
- 04 **EDA on important features**
Here we analyze the data
- 05 **Conclusion**
Recommendation about project

Introduction



- Coupon systems have been widely used to market products, and services and engage customers to use their products and services often. Coupons create a win-win situation for both companies and customers so, by offering a correct coupon to users, which can lead users to become frequent customers and it is enhancing a brand's impact on its customers.
- How to know which coupon to provide a customer can be a rather complex task, since each customer profile responds differently to each other, and frequently offering them bad coupons or deals might drag them away from our business. To overcome this problem, machine learning techniques can be used to build a better coupon recommendations system.
- Predicting whether the customer will accept the coupon or not is a difficult problem and we can not just recommend it to everyone because of the costs concerned so, in this problem, we will predict whether a customer will accept or reject the offered coupon based on the customer's profile and history.
- This prediction helps the company in offering a correct coupon so that more customers will use their services which leads to more business for the company.

Executive Summary

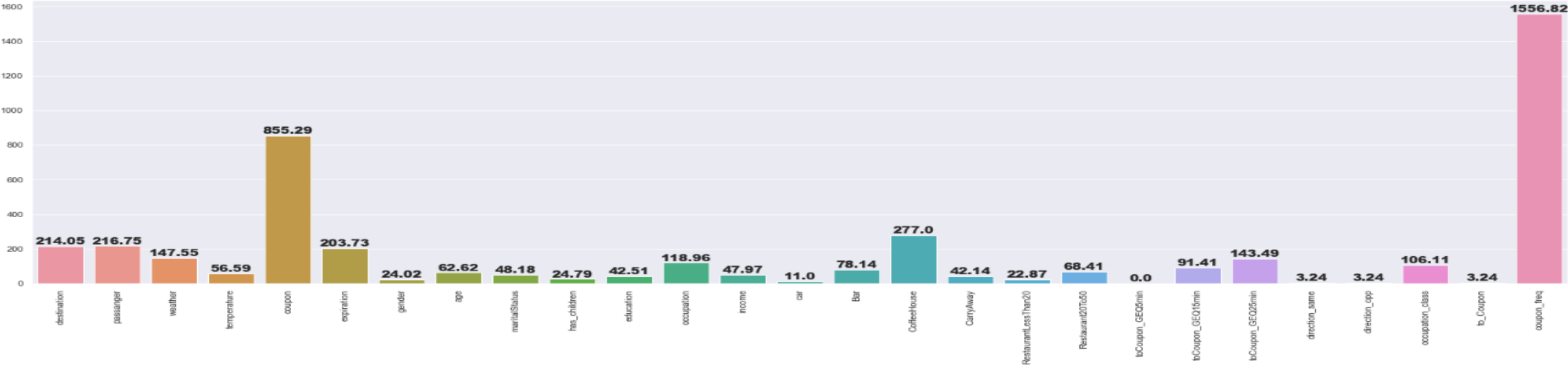
- ▶ We have done significant amount of data cleaning and feature engineering in this dataset.
- ▶ There are 3 features which are toCoupon_GEQ5min, direction_same, direction_opp are failed in Hypothesis test and car has more than 99% missing values, we drop those variables.
- ▶ By combining toCoupon_GEQ15min and toCoupon_GEQ25min feature we created another variable to_coupon and drop those 2 variables.
- ▶ Similarly, we created another 2 variables **coupon_freq** and **occupation_class** by doing feature engineering.
- ▶ Xgboost is our optimum model for classification of coupon acceptance or rejection.
- ▶ As this is model for e-commerce domain, here precision is most important evaluation metrics as we are more concerned about positives.
- ▶ Although Adaboost also gives same precision, but it have significant amount of variance.

Hypothesis Test Result

	column	chisq_stat	pval
0	destination	214.045718	0.000000
1	passanger	216.745407	0.000000
2	weather	147.554996	0.000000
3	temperature	56.586304	0.000000
4	coupon	855.294266	0.000000
5	expiration	203.729115	0.000000
6	gender	24.023679	0.000001
7	age	62.615040	0.000000
8	maritalStatus	48.175873	0.000000
9	has_children	24.788591	0.000001
10	education	42.506111	0.000000
11	occupation	118.955538	0.000000
12	income	47.966117	0.000000
13	car	11.000918	0.026554

14	Bar	78.135330	0.000000
15	CoffeeHouse	277.003931	0.000000
16	CarryAway	42.137920	0.000000
17	RestaurantLessThan20	22.874095	0.000134
18	Restaurant20To50	68.413594	0.000000
19	toCoupon_GEQ5min	0.000000	1.000000
20	toCoupon_GEQ15min	91.414686	0.000000
21	toCoupon_GEQ25min	143.489554	0.000000
22	direction_same	3.235334	0.072066
23	direction_opp	3.235334	0.072066
24	Accept(Y/N?)	12388.929832	0.000000
25	occupation_class	106.110860	0.000000
26	to_Coupon	180.589654	0.000000
27	coupon_freq	1556.815739	0.000000
28	feature_comb	208.155590	0.000000

Chi2 Statistic Value of each Categorical Columns

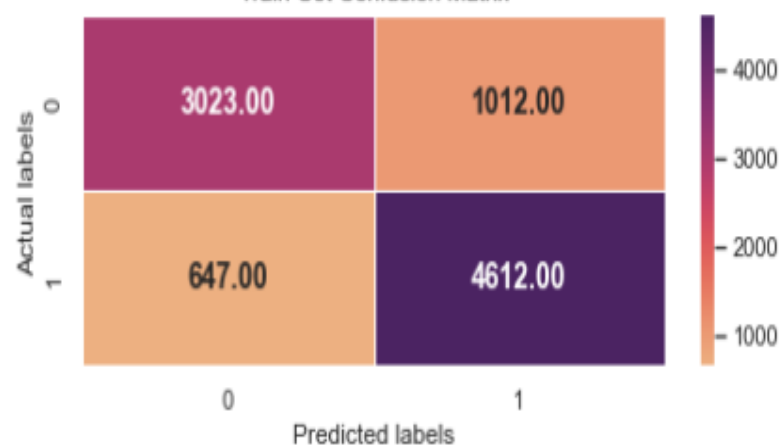


XGBClassifier model evaluation metrics

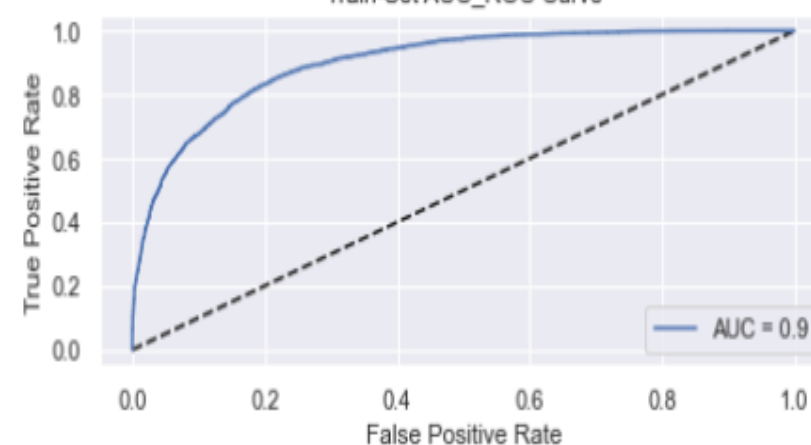
Train-Set Classification Report



Train-Set Confusion Matrix



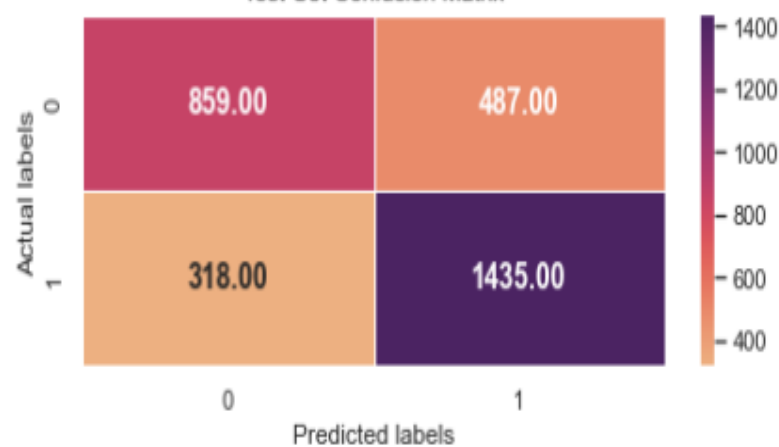
Train-Set AUC_ROC Curve



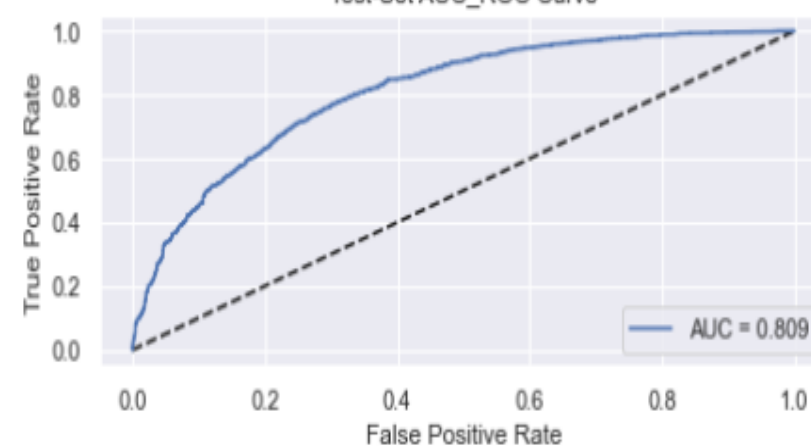
Test-Set Classification Report



Test-Set Confusion Matrix

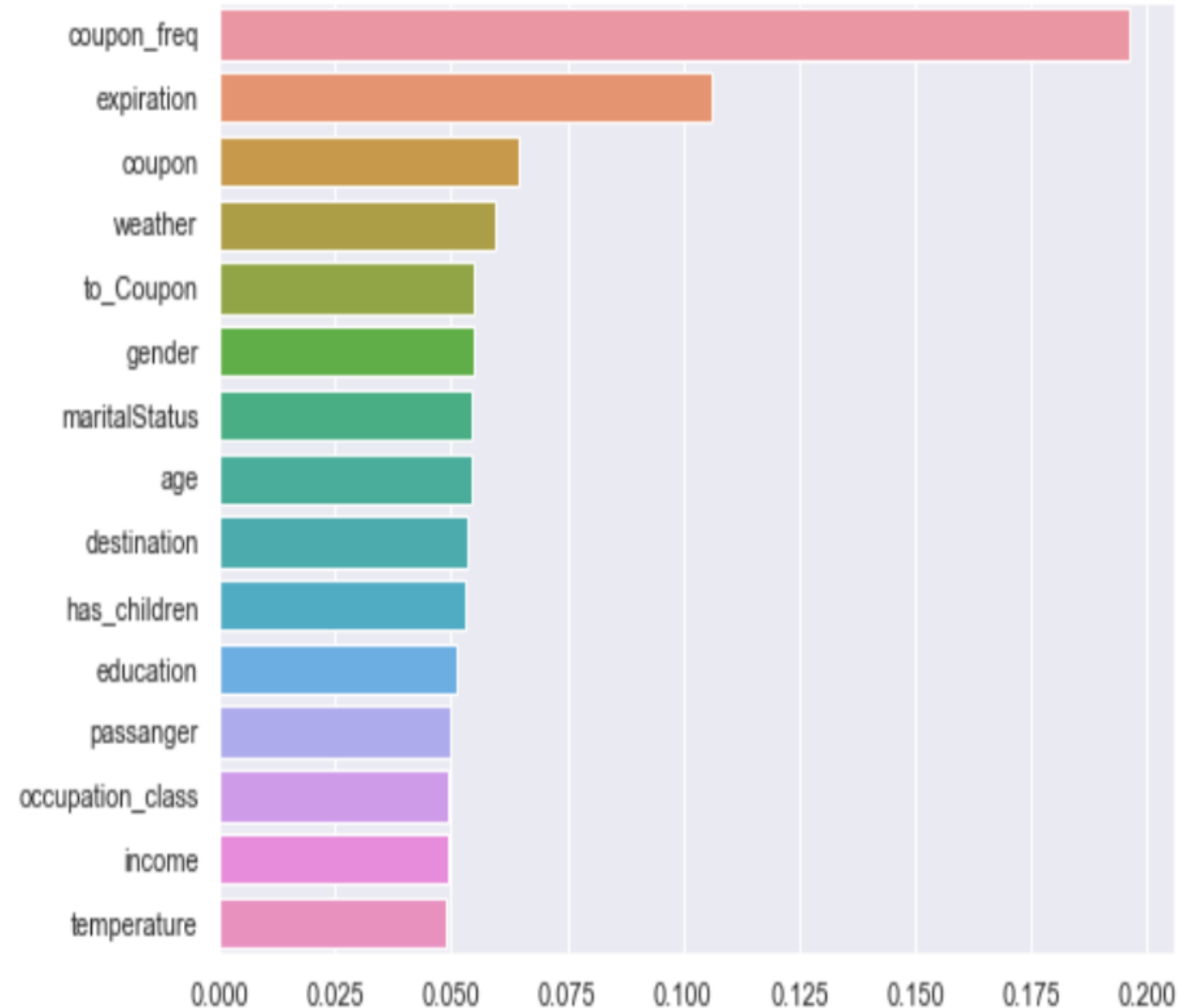


Test-Set AUC_ROC Curve

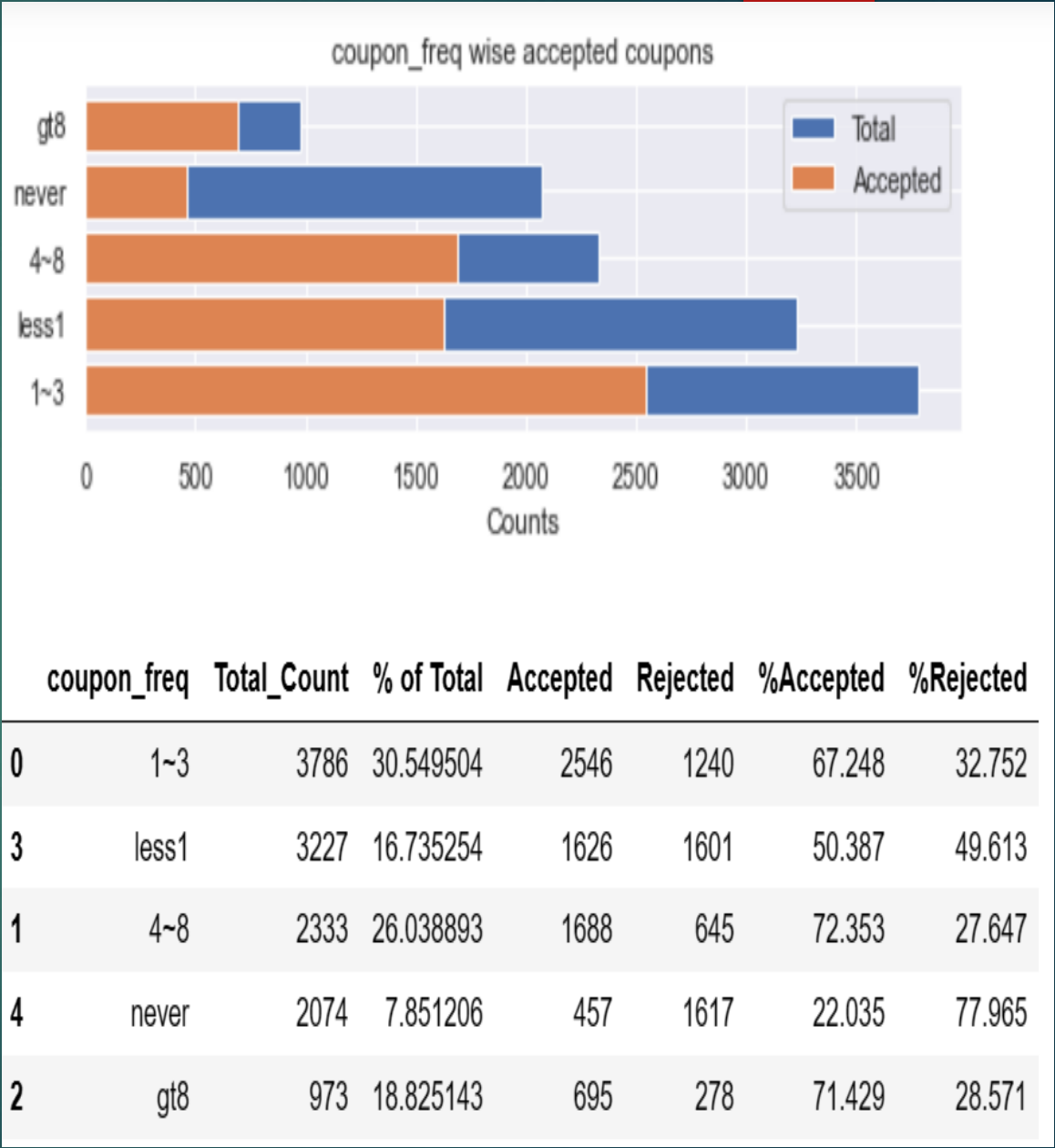
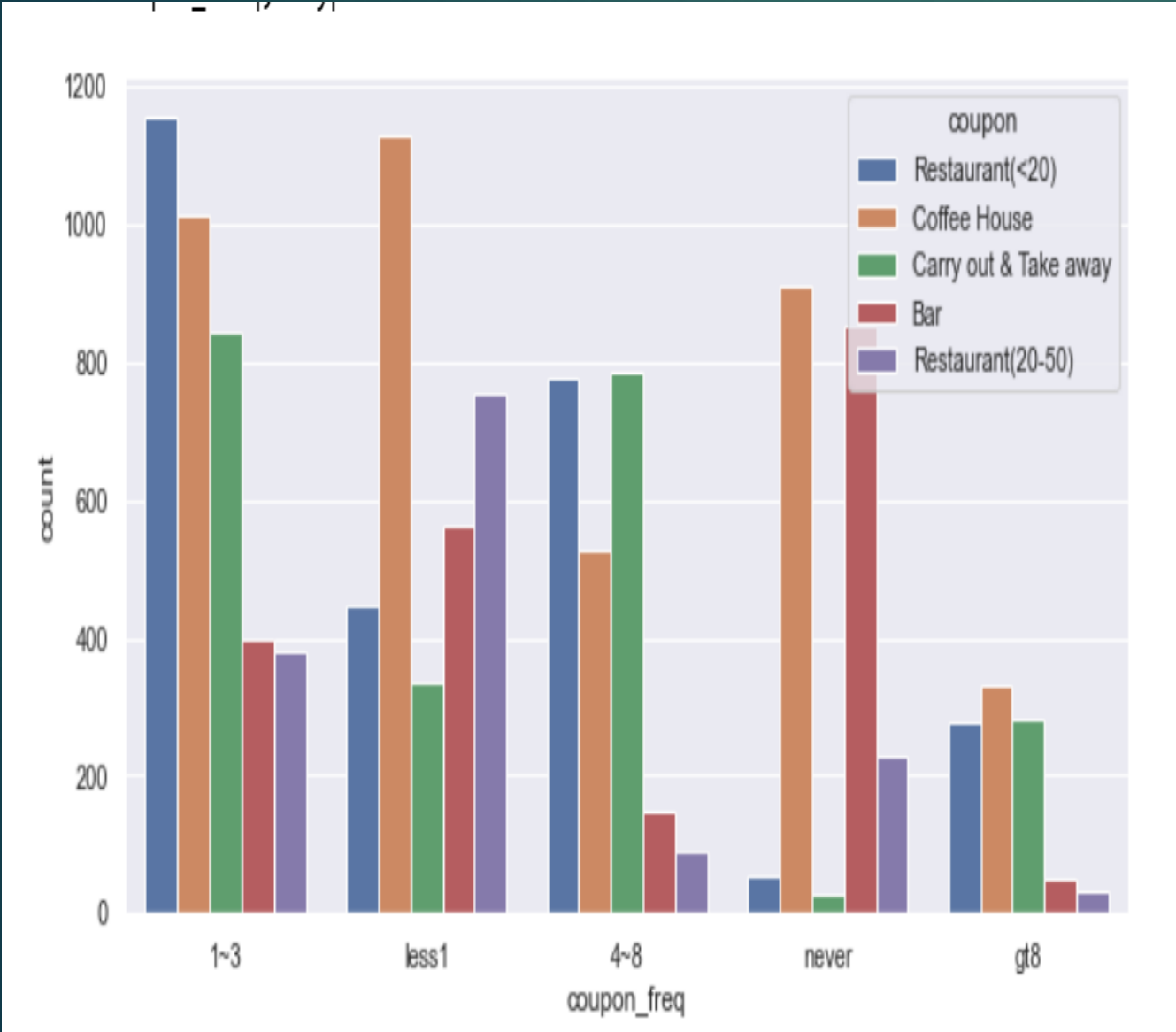


Feature Importance and all model results

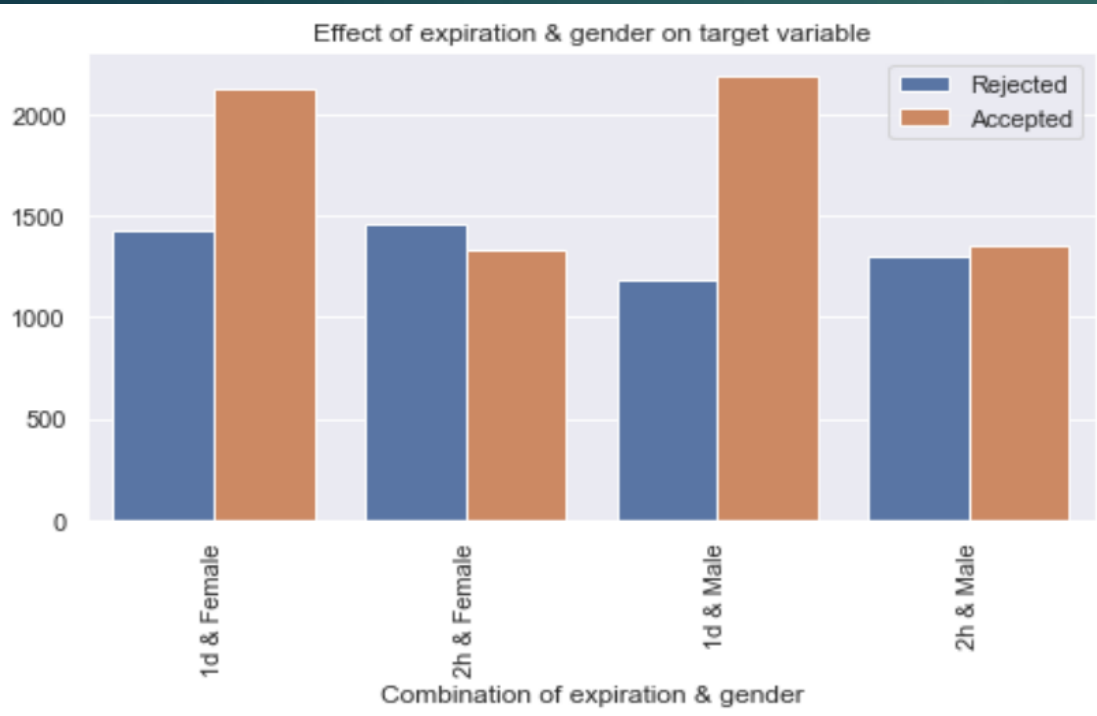
	Models	Precision	Recall	F1_Score	Accuracy	AUC
0	Logistic_regression	0.70	0.69	0.69	0.70	0.764
1	Decision_tree	0.66	66.00	0.66	0.66	0.665
2	Random_forest	0.73	0.72	0.72	0.73	0.789
3	AdaBoost	0.74	0.72	0.73	0.74	0.768
4	XGboost	0.74	0.73	0.73	0.74	0.809
5	SVM	0.72	0.70	0.70	0.72	0.784



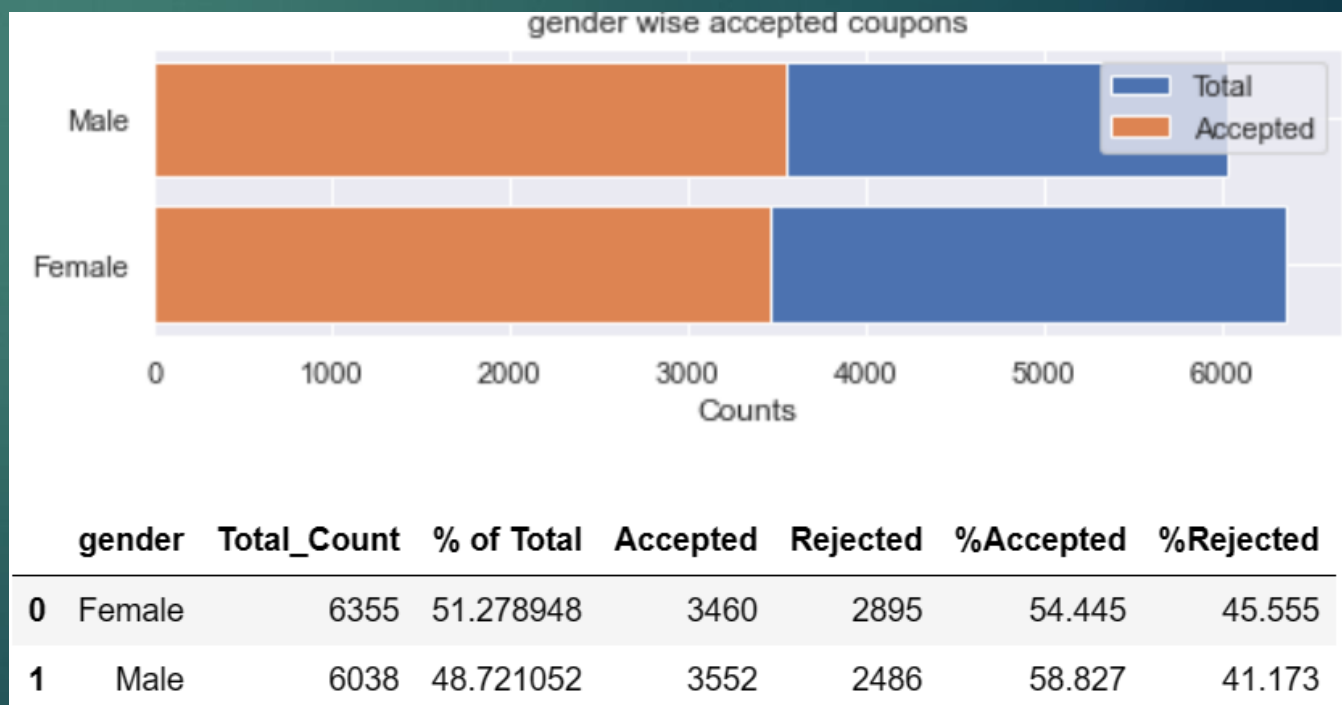
coupon_freq



Expiration and gender

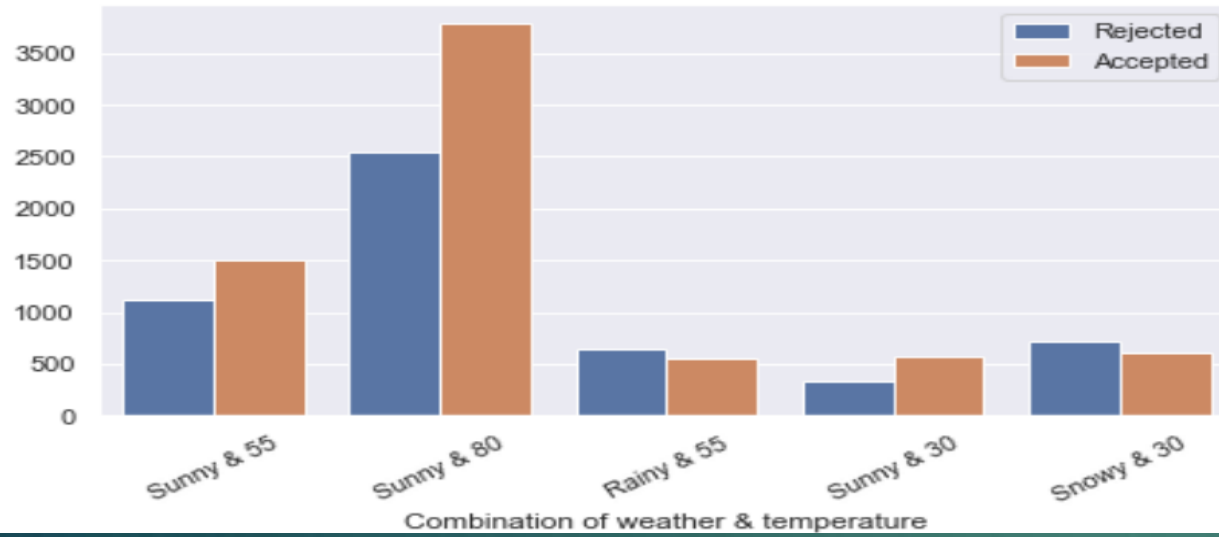


- When validity of coupon is 1d, more people accept and when validity is 2h the acceptance ratio is low.
- Although more female participate in survey, still coupon acceptance ratio is better in males.
- Both Male and Female participants accepts coupon with validity of 1d, but more females reject coupons with validity of 2h.

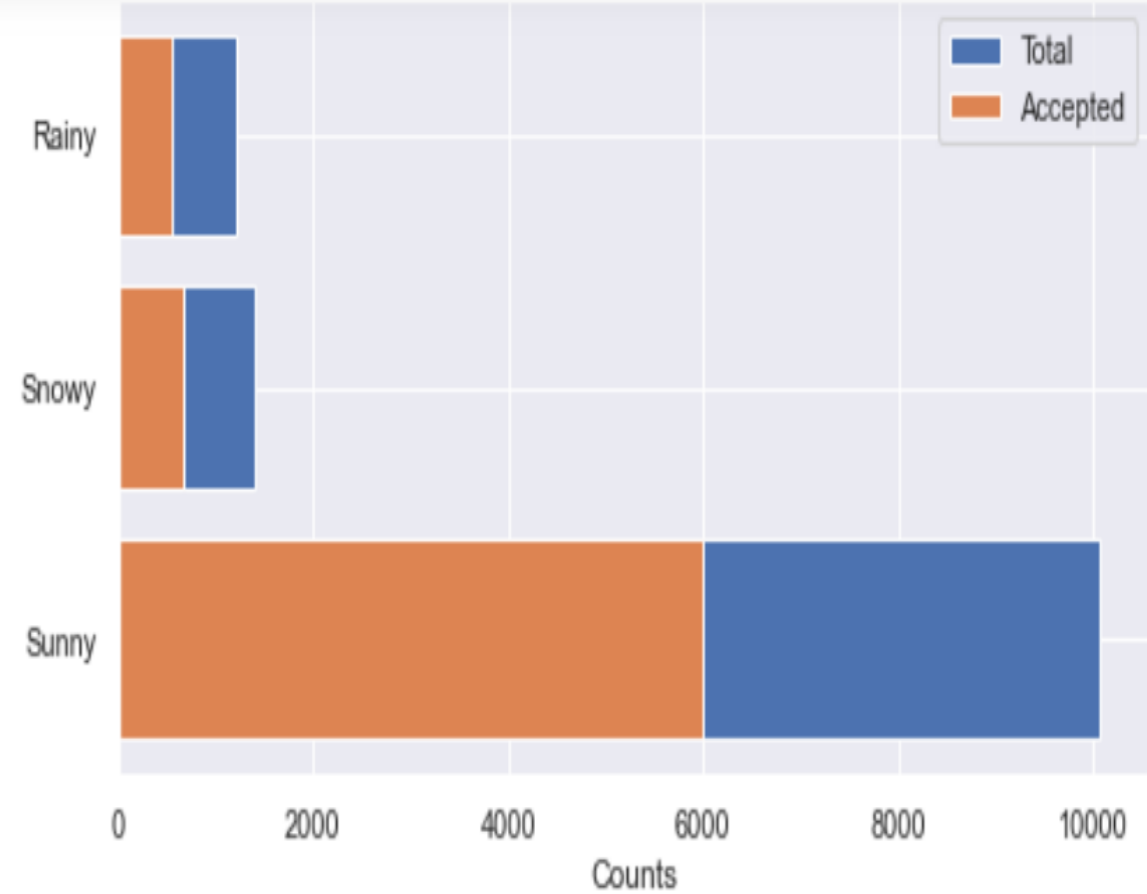
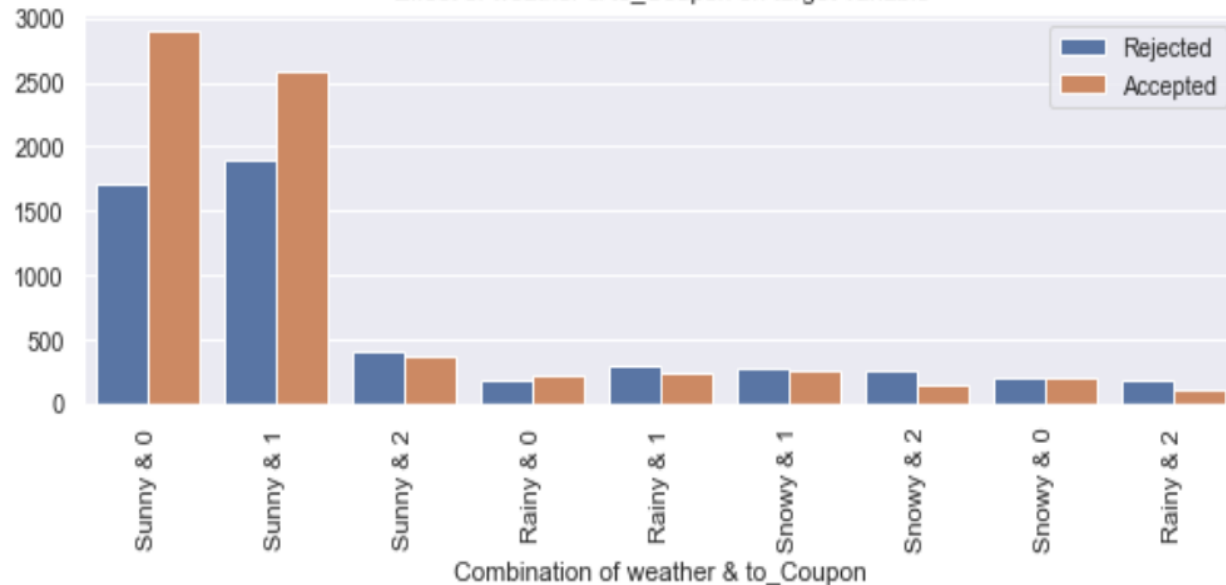


weather

Effect of weather & temperature on target variable

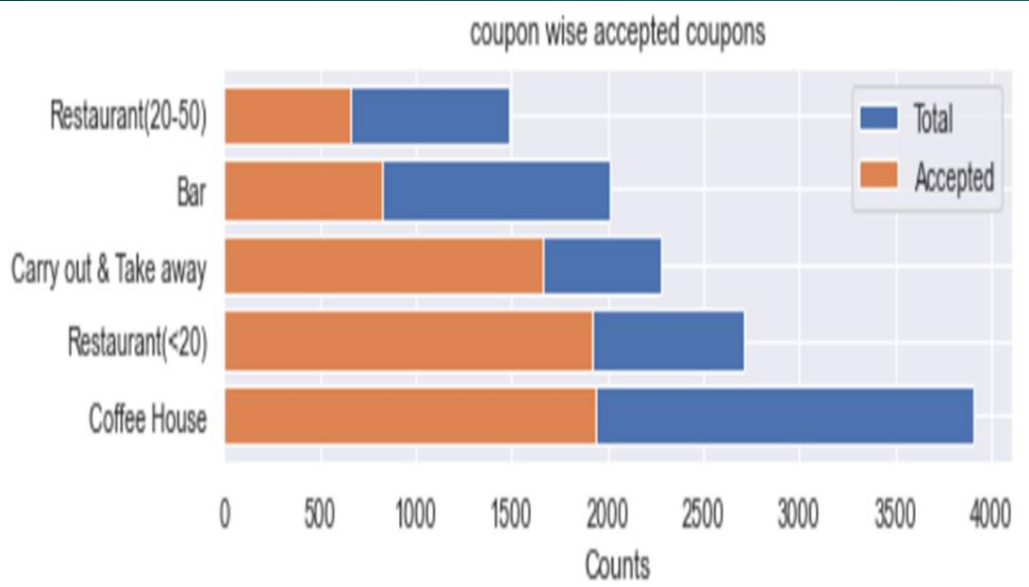


Effect of weather & to_Coupon on target variable

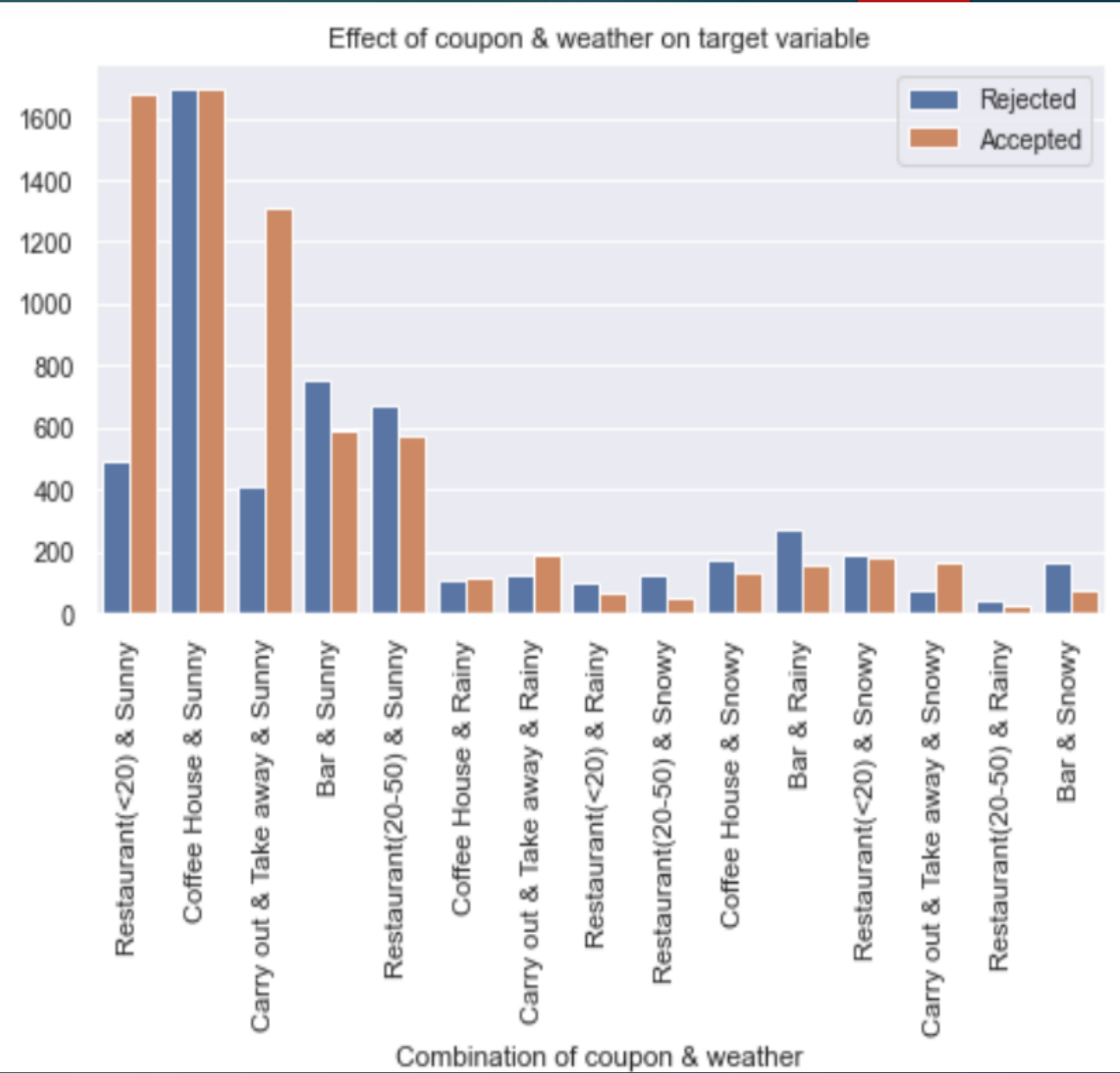


	weather	Total_Count	% of Total	Accepted	Rejected	%Accepted	%Rejected
2	Sunny	10069	9.539577	5989	4080	59.480	40.520
1	Snowy	1405	11.076947	661	744	47.046	52.954
0	Rainy	1210	79.383475	560	650	46.281	53.719

coupon



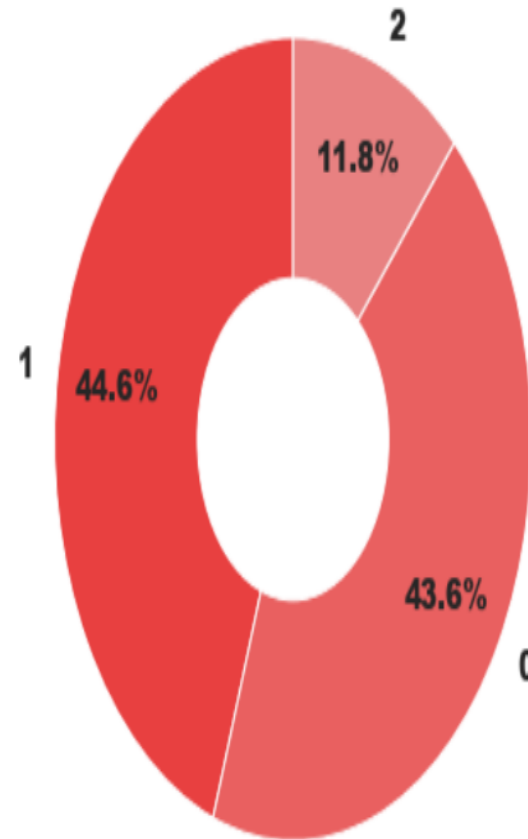
	coupon	Total_Count	% of Total	Accepted	Rejected	%Accepted	%Rejected
2	Coffee House	3913	18.357137	1943	1970	49.655	50.345
4	Restaurant(<20)	2708	11.998709	1924	784	71.049	28.951
1	Carry out & Take away	2275	21.851045	1665	610	73.187	26.813
0	Bar	2010	31.574276	824	1186	40.995	59.005
3	Restaurant(20-50)	1487	16.218833	656	831	44.116	55.884



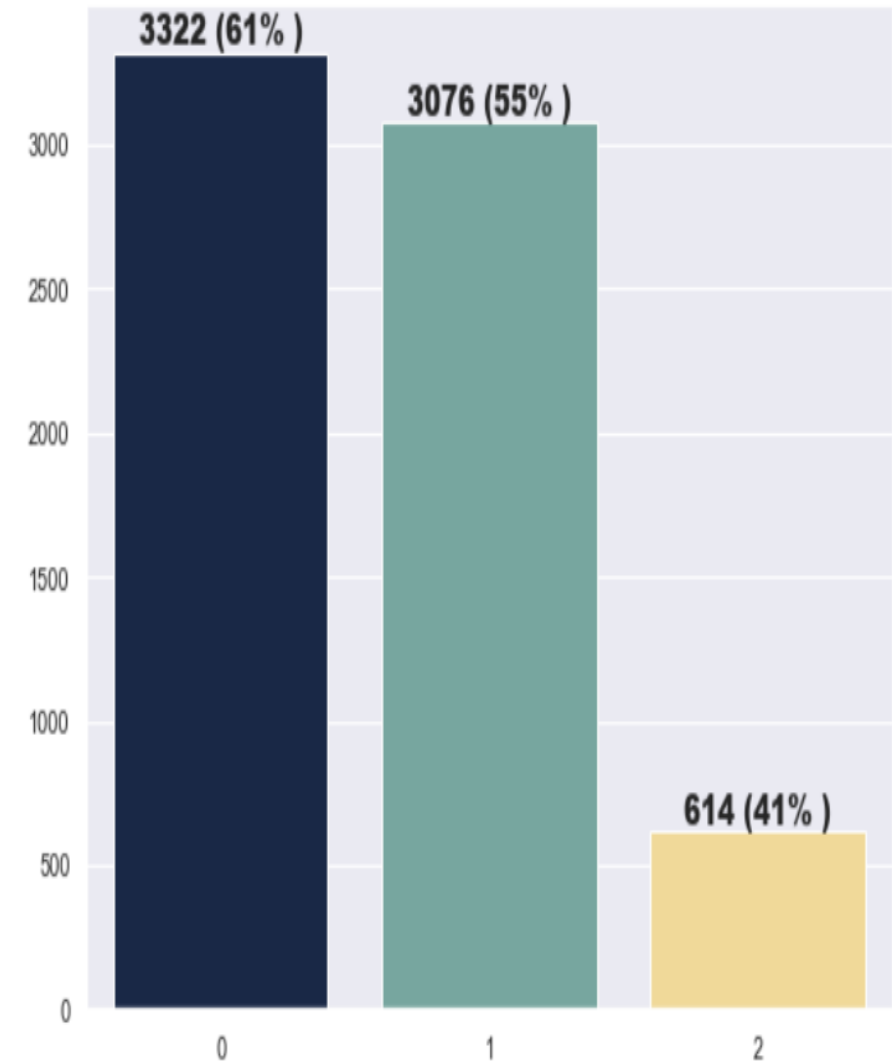
to_coupon

- ▶ to_coupon column is combination of to_Coupon_GEQ15min and to_Coupon_GEQ25min is done through feature engineering.
- ▶ This feature shows people whose distance is within 15minutes are more likely to accept coupon than within15mins and within25mins.

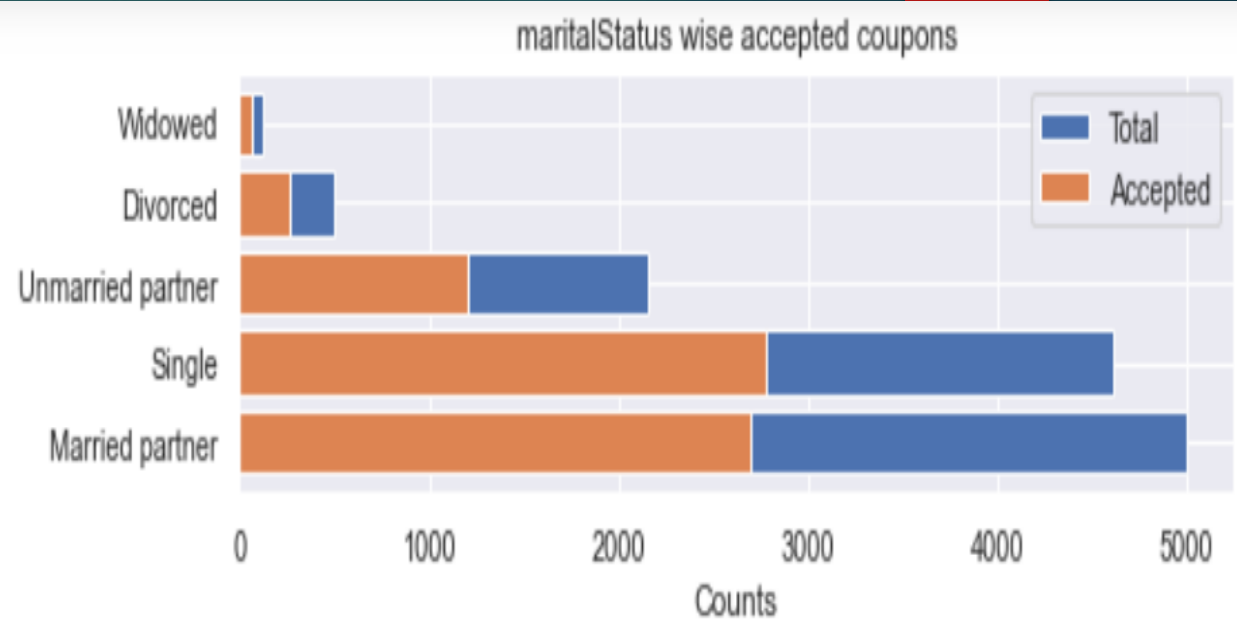
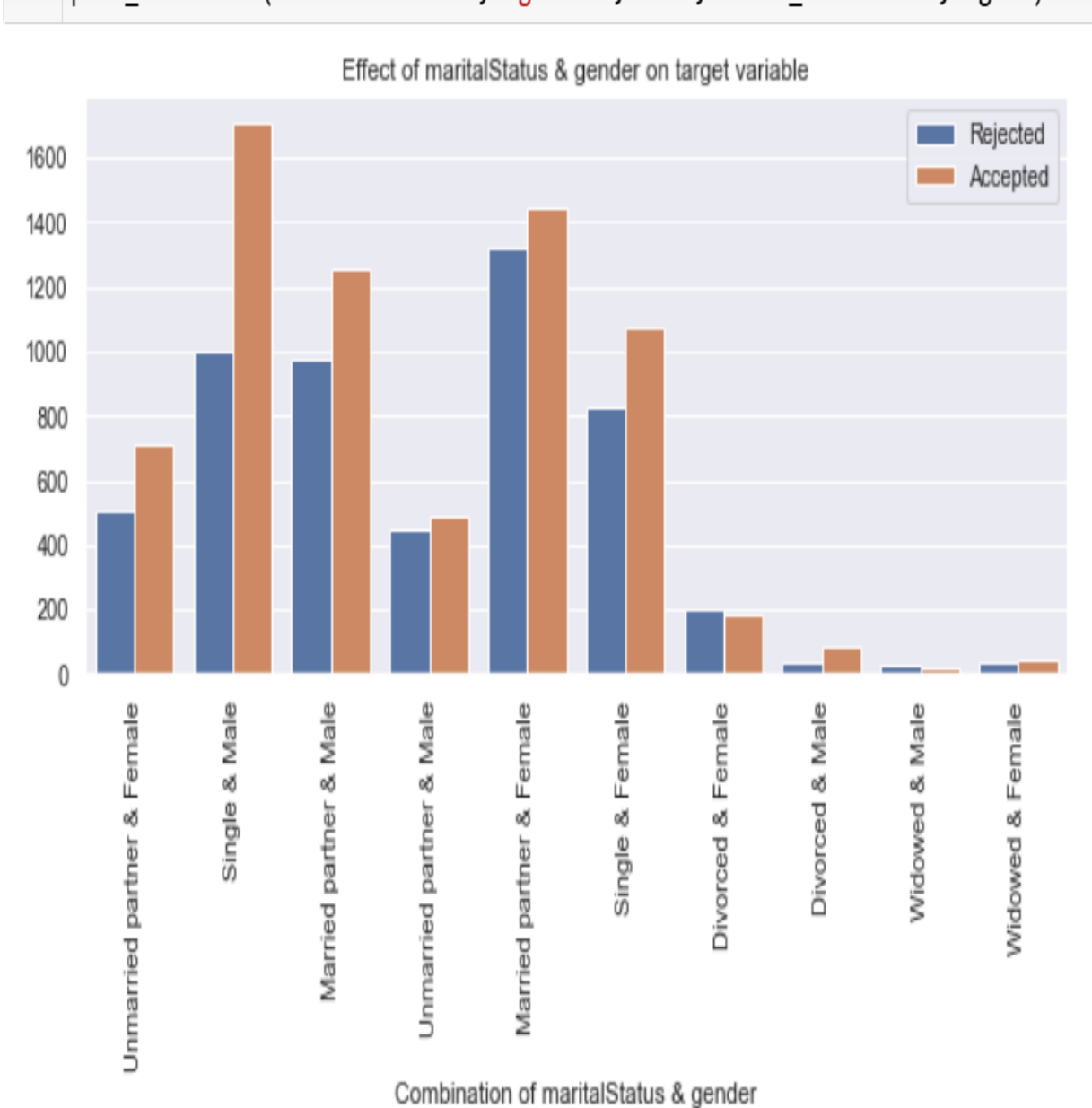
Coupon Acceptance by to_Coupon



Coupon Acceptance by to_Coupon

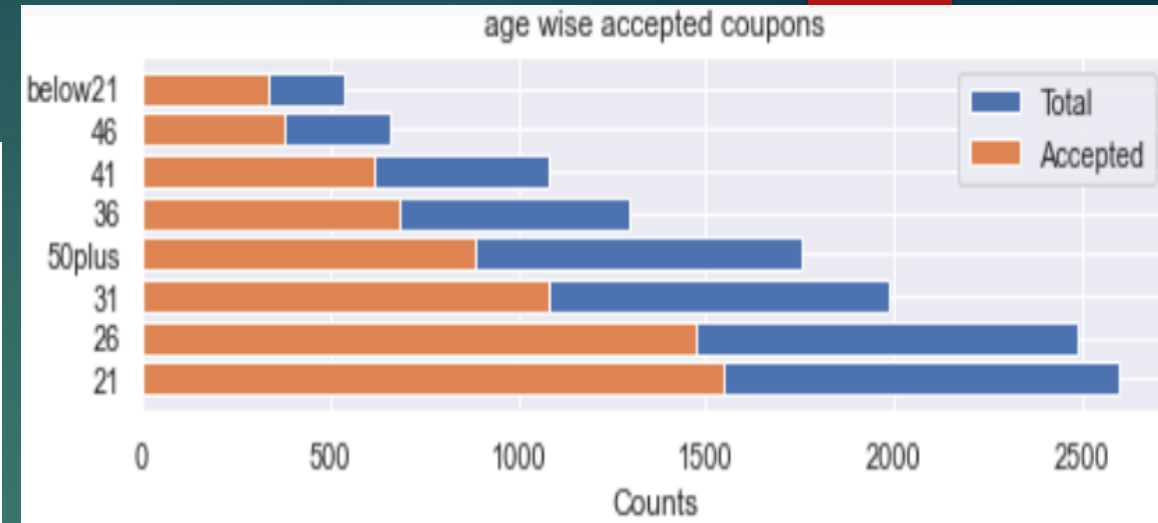
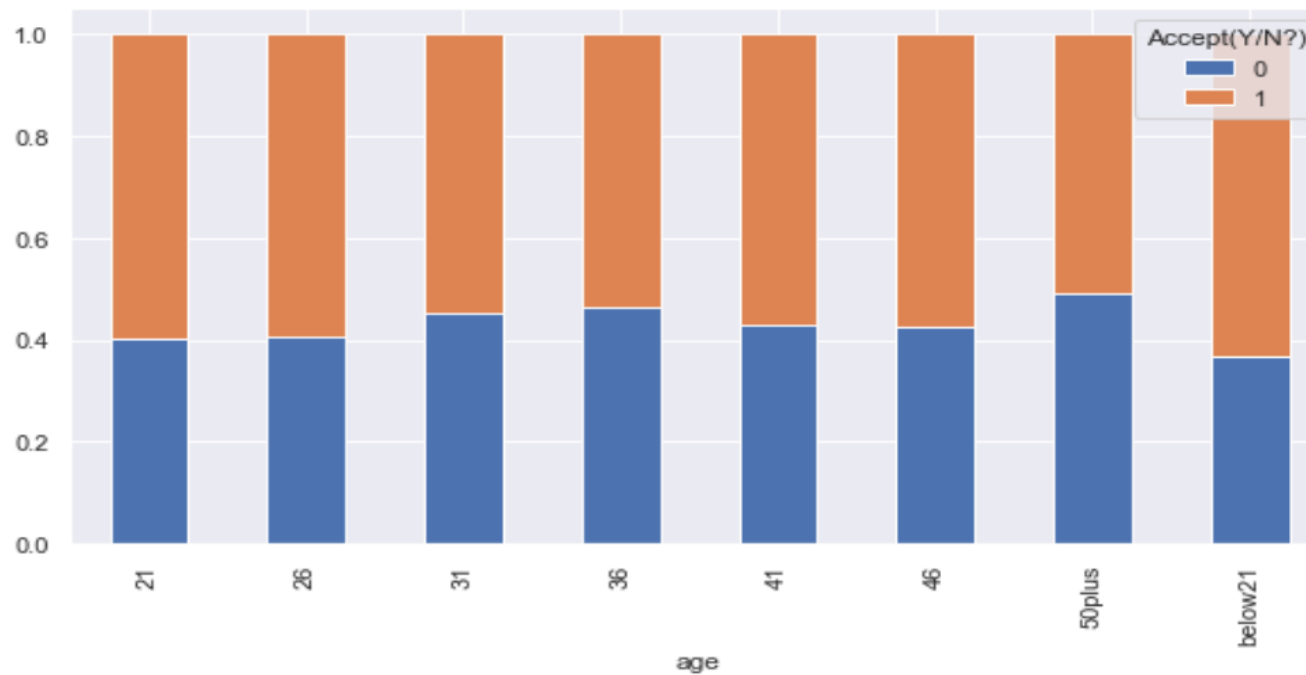
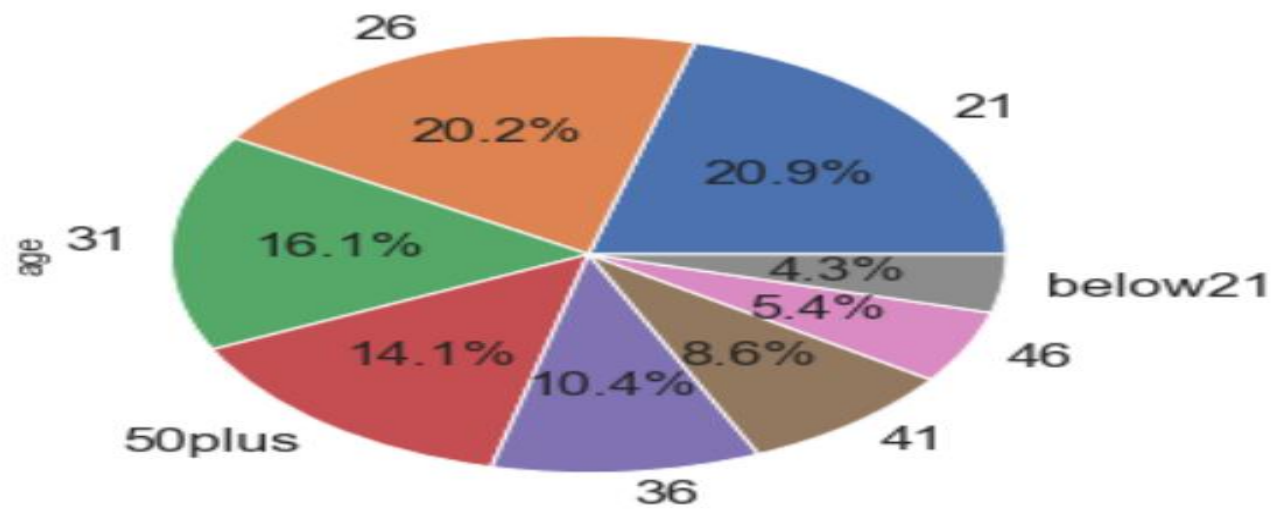


maritalStatus



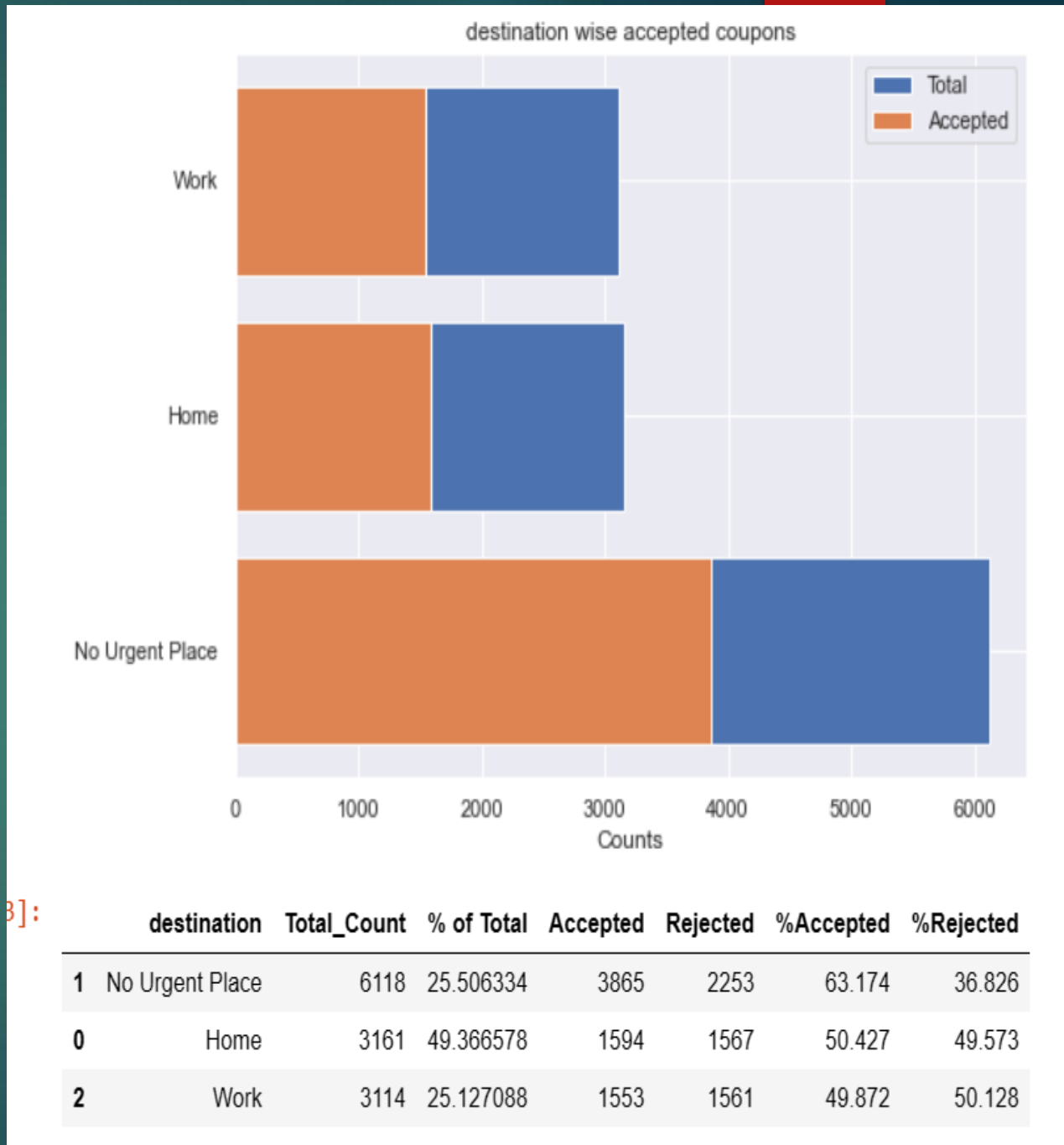
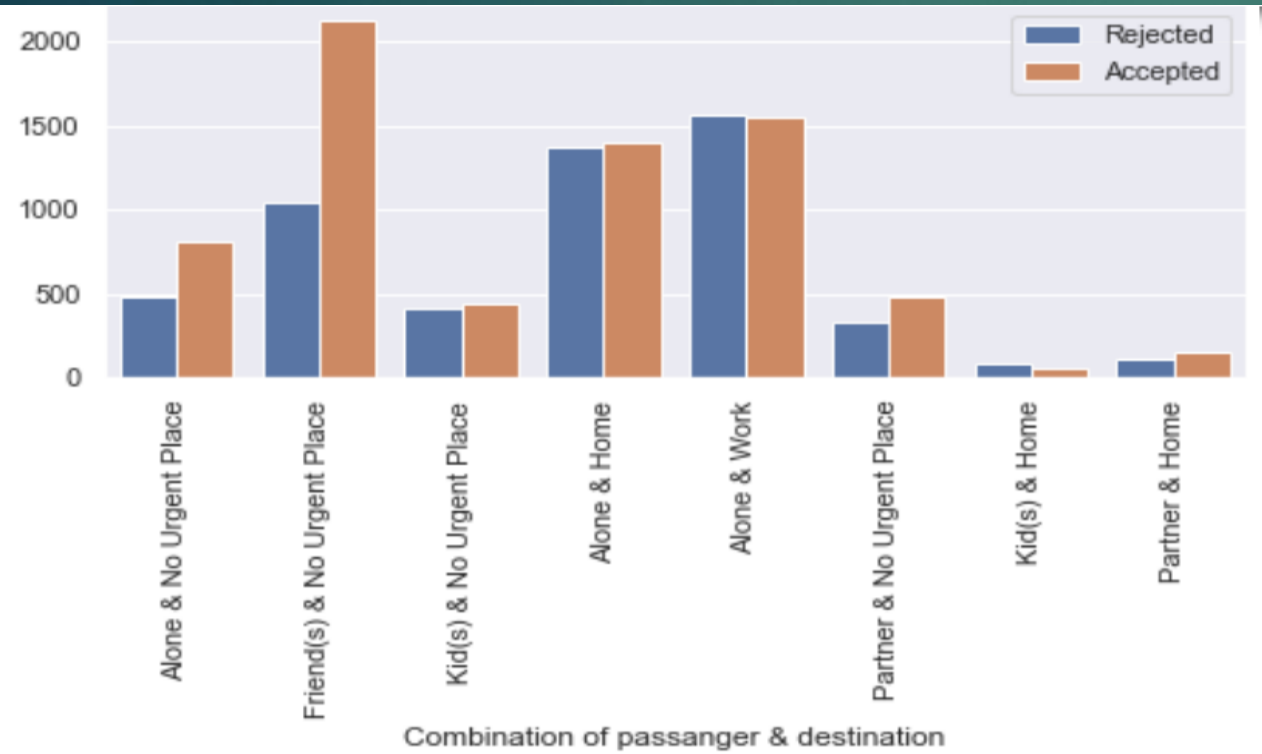
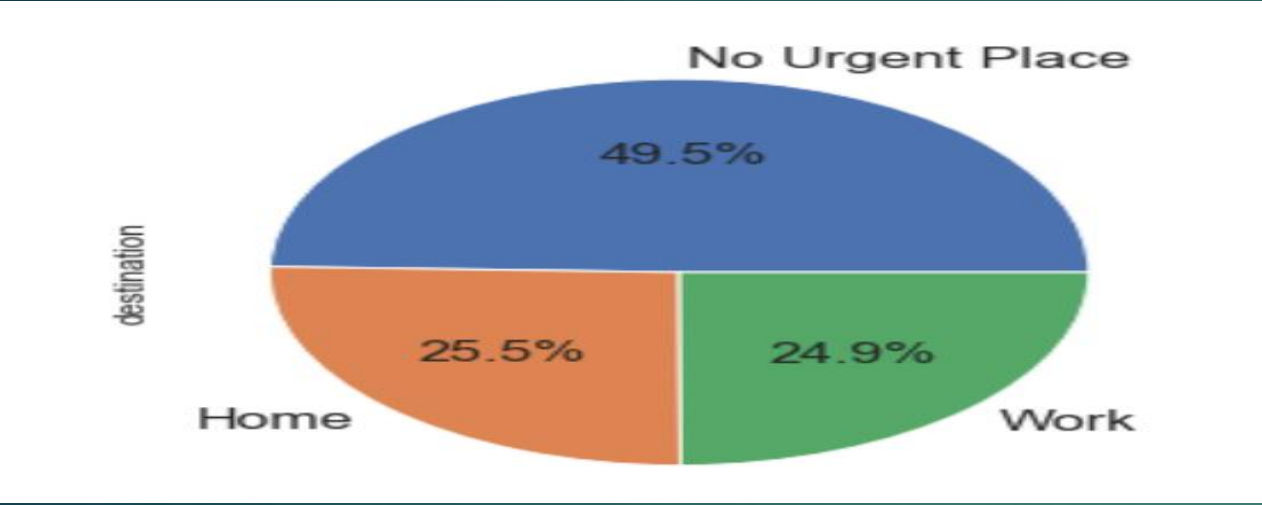
	maritalStatus	Total_Count	% of Total	Accepted	Rejected	%Accepted	%Rejected
1	Married partner	4994	37.198418	2697	2297	54.005	45.995
2	Single	4610	17.388849	2784	1826	60.390	39.610
3	Unmarried partner	2155	4.074881	1203	952	55.824	44.176
0	Divorced	505	40.296942	266	239	52.673	47.327
4	Widowed	129	1.040910	62	67	48.062	51.938

age



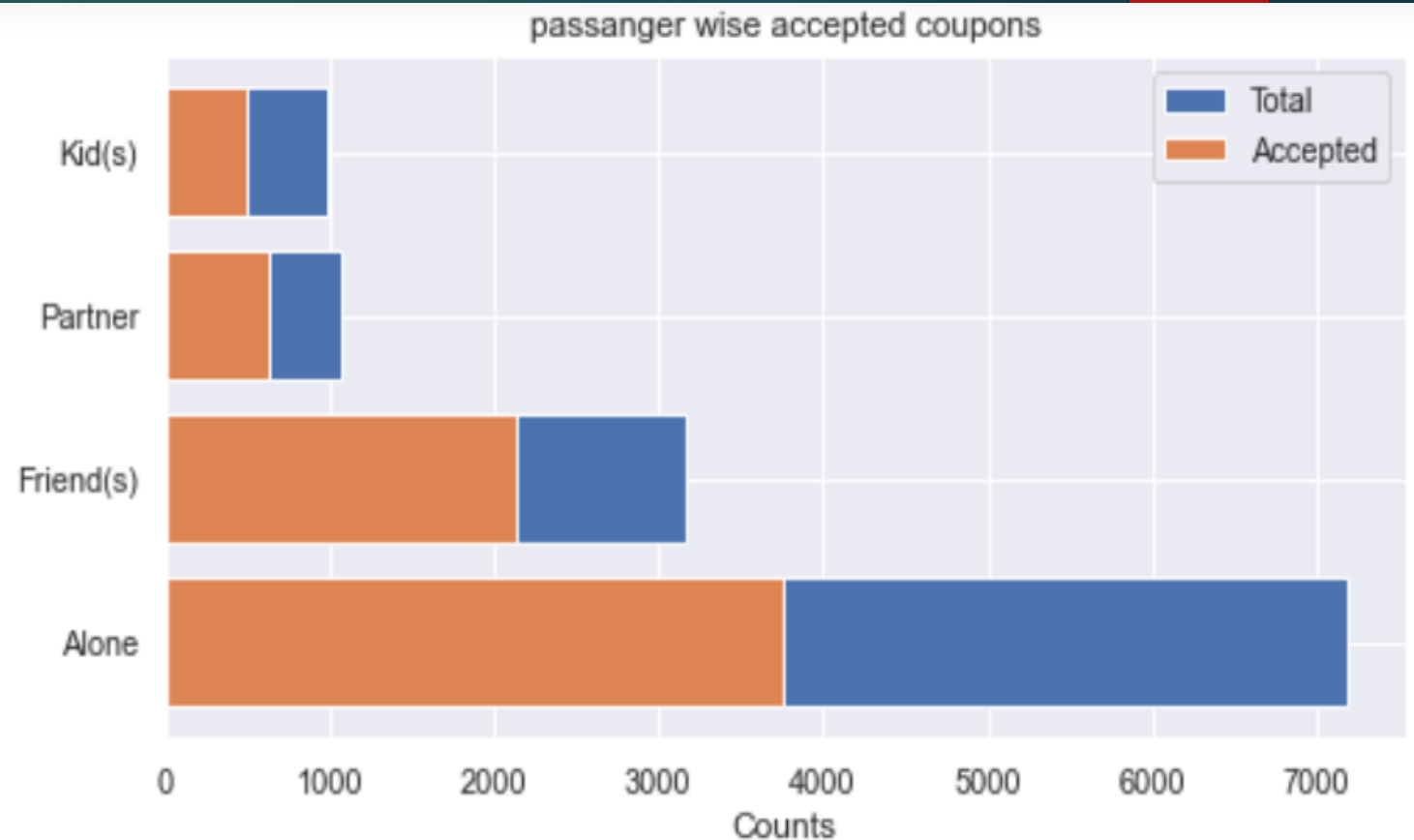
	age	Total_Count	% of Total	Accepted	Rejected	%Accepted	%Rejected
0	21	2598	20.963447	1549	1049	59.623	40.377
1	26	2486	20.059711	1475	1011	59.332	40.668
2	31	1984	16.009037	1079	905	54.385	45.615
6	50plus	1756	5.325587	888	868	50.569	49.431
3	36	1293	14.169289	687	606	53.132	46.868
4	41	1079	10.433309	618	461	57.275	42.725
5	46	660	8.706528	378	282	57.273	42.727
7	below21	537	4.333091	338	199	62.942	37.058

Destination



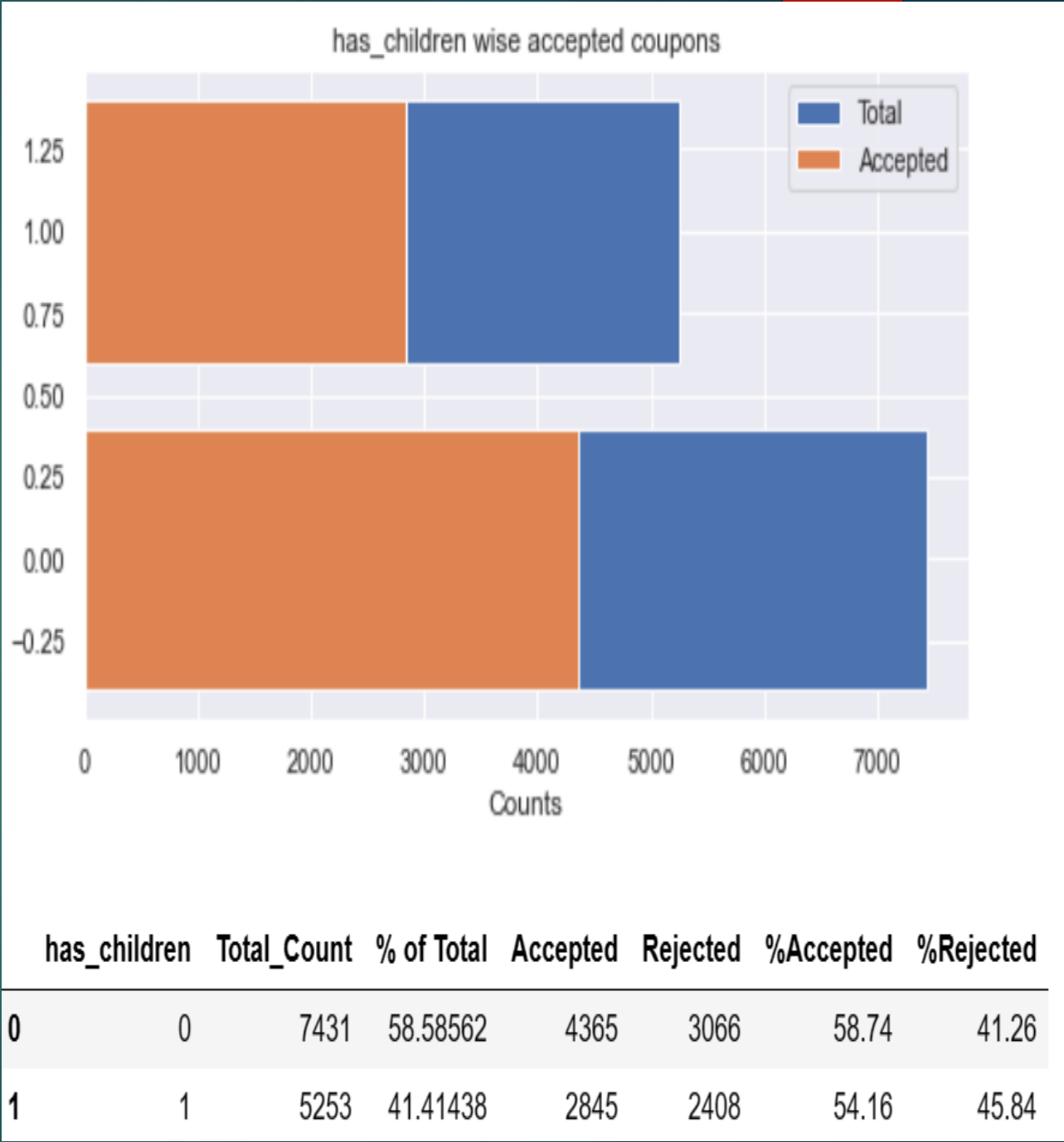
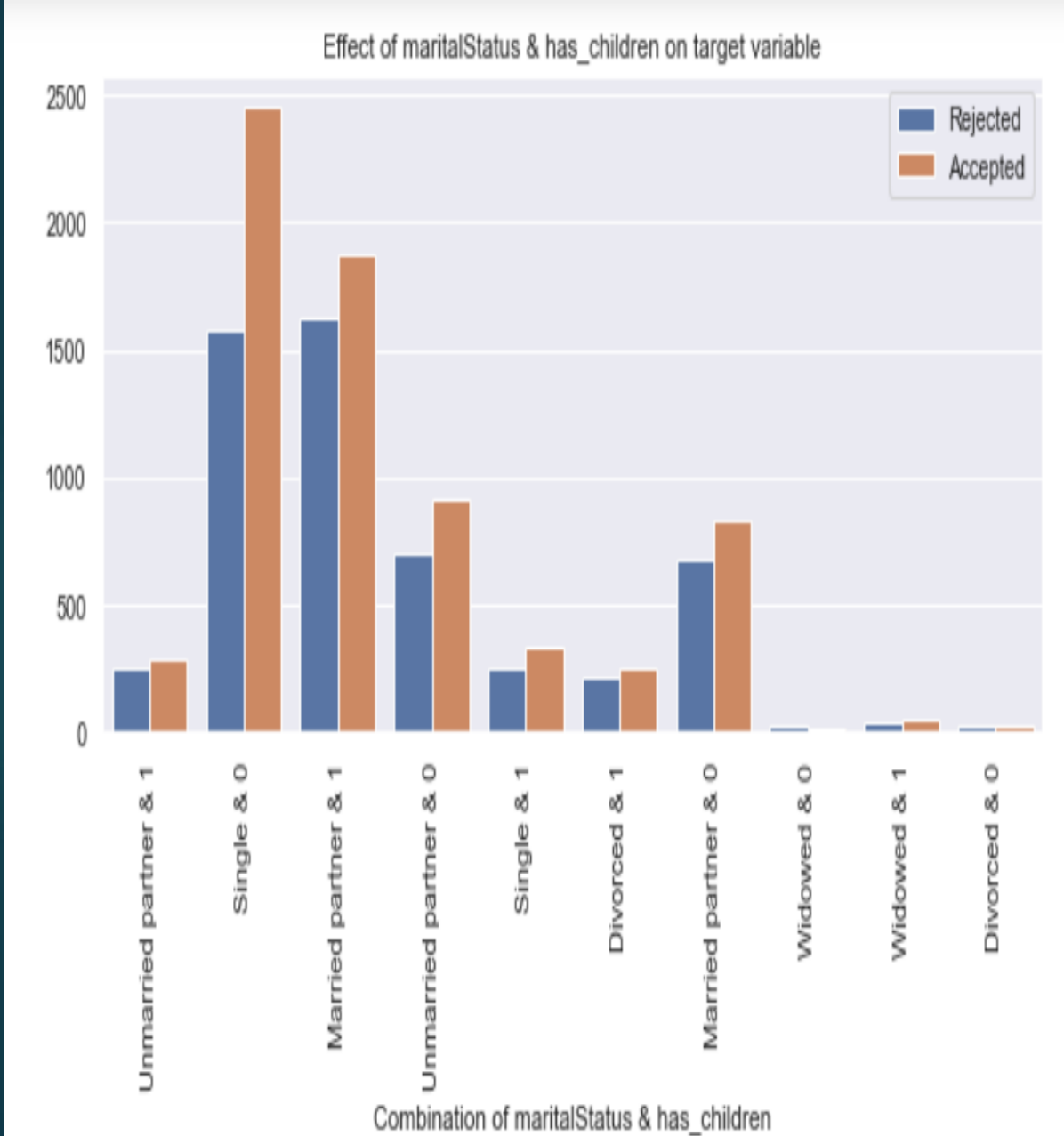
passanger

- ▶ Passanger who are alone have attended the survey in more numbers.
- ▶ Passanger with Friend(s) and Partner more likely to accept the coupons than others. So, we need to focus on them.
- ▶ Also we need to focus on Alone passangers who participate in huge numbers almost 58%, but their acxceptance ratio is low.
- ▶ Also need to focus on passanger with kids, who have less acceptance ratio among all groups, may be we need to have something which will beneficial for kid(s).

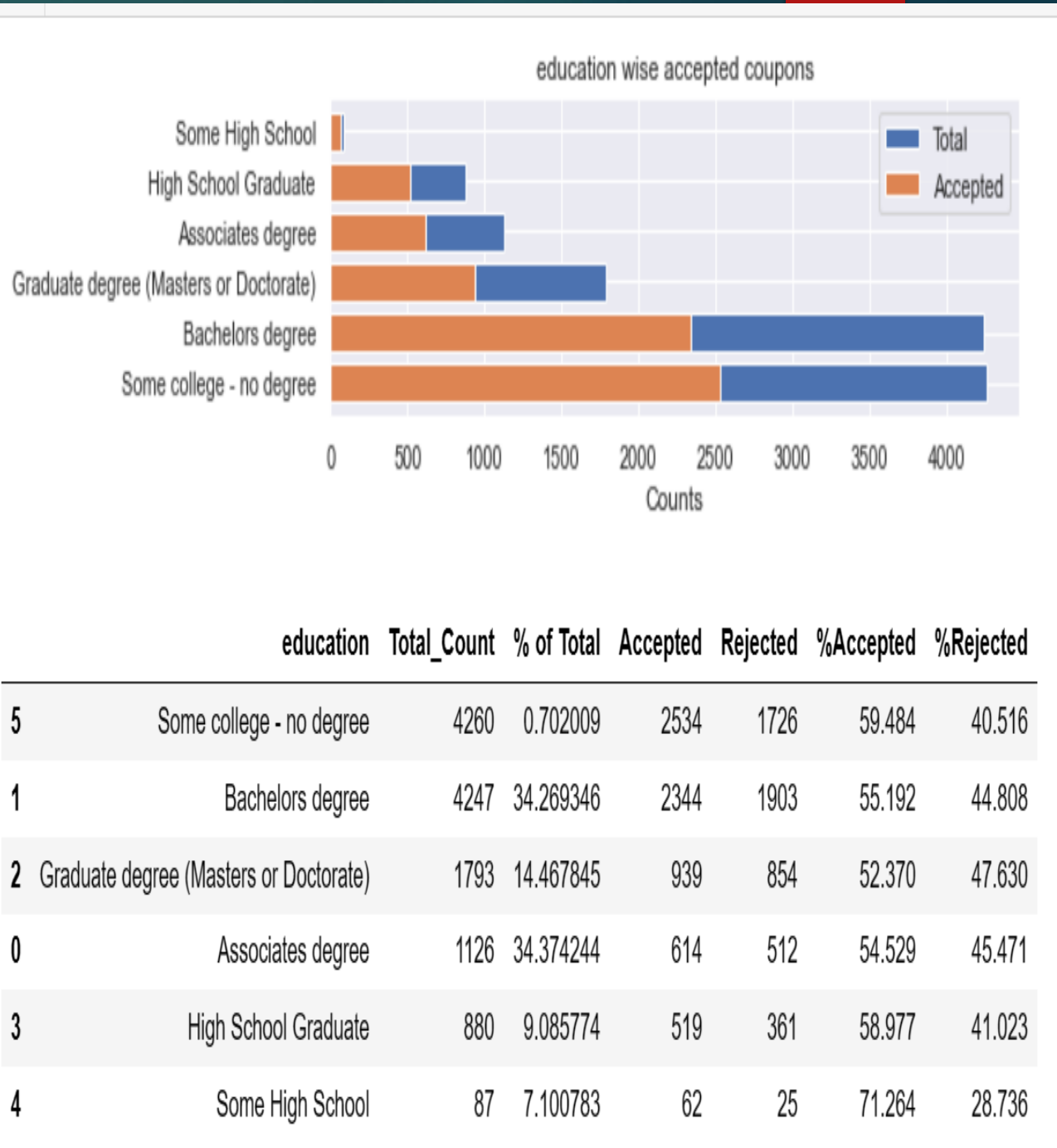
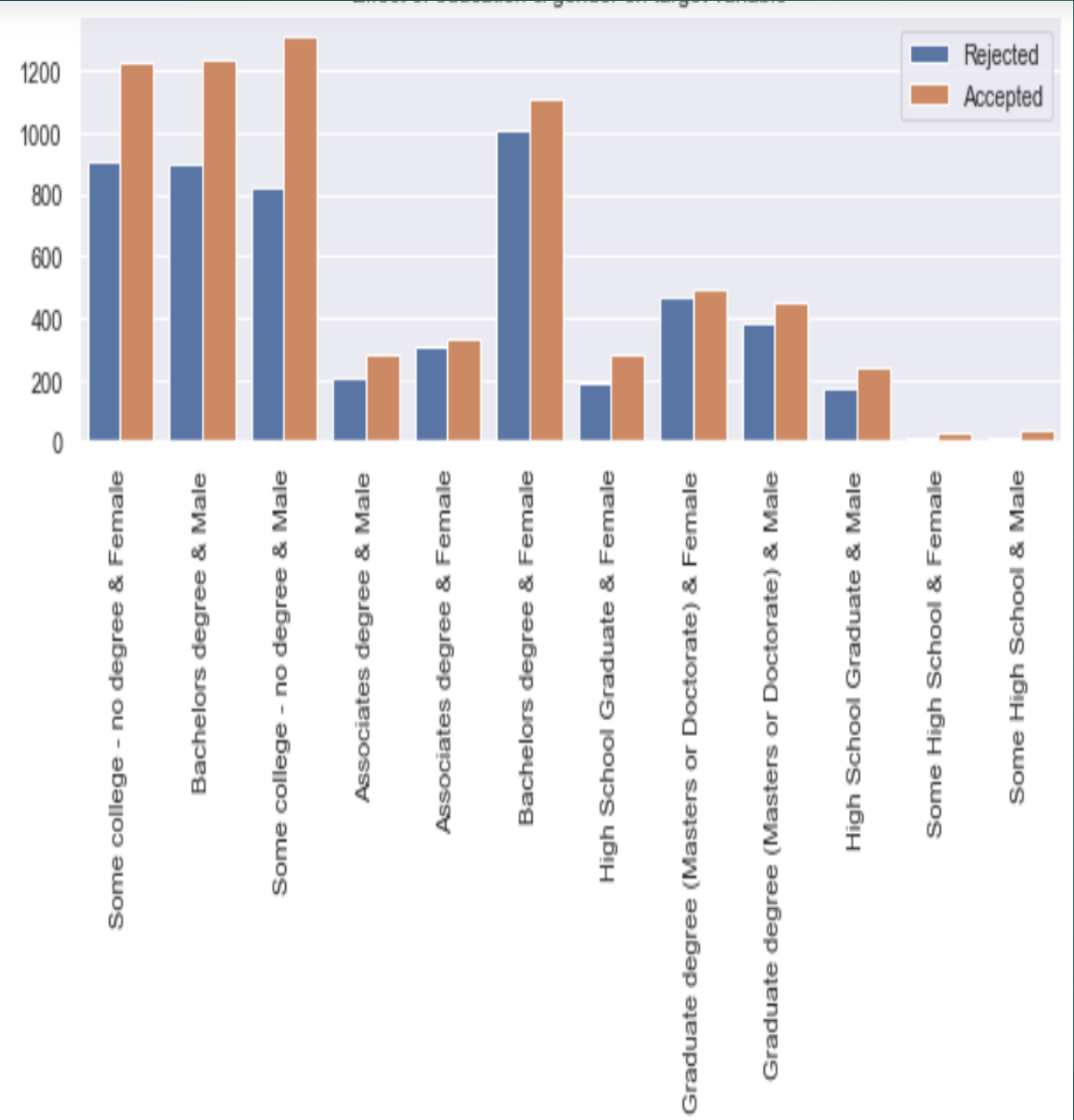


	passanger	Total_Count	% of Total	Accepted	Rejected	%Accepted	%Rejected
0	Alone	7178	57.919793	3760	3418	52.382	47.618
1	Friend(s)	3166	25.546680	2127	1039	67.183	32.817
3	Partner	1068	7.915759	634	434	59.363	40.637
2	Kid(s)	981	8.617768	491	490	50.051	49.949

has_children

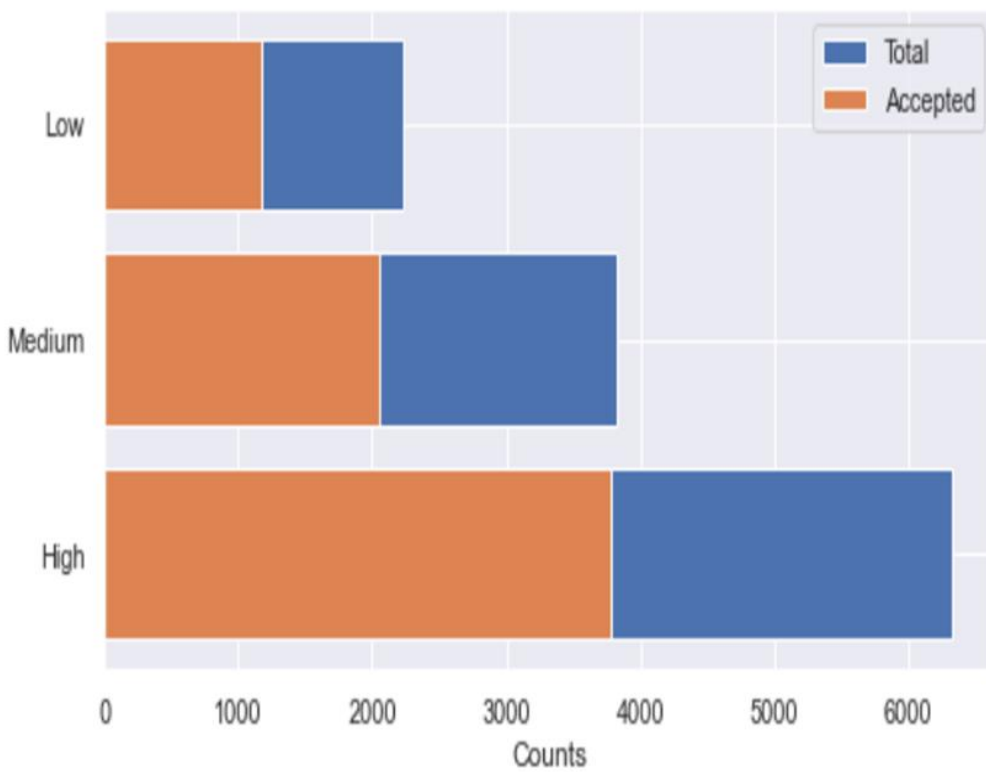


education



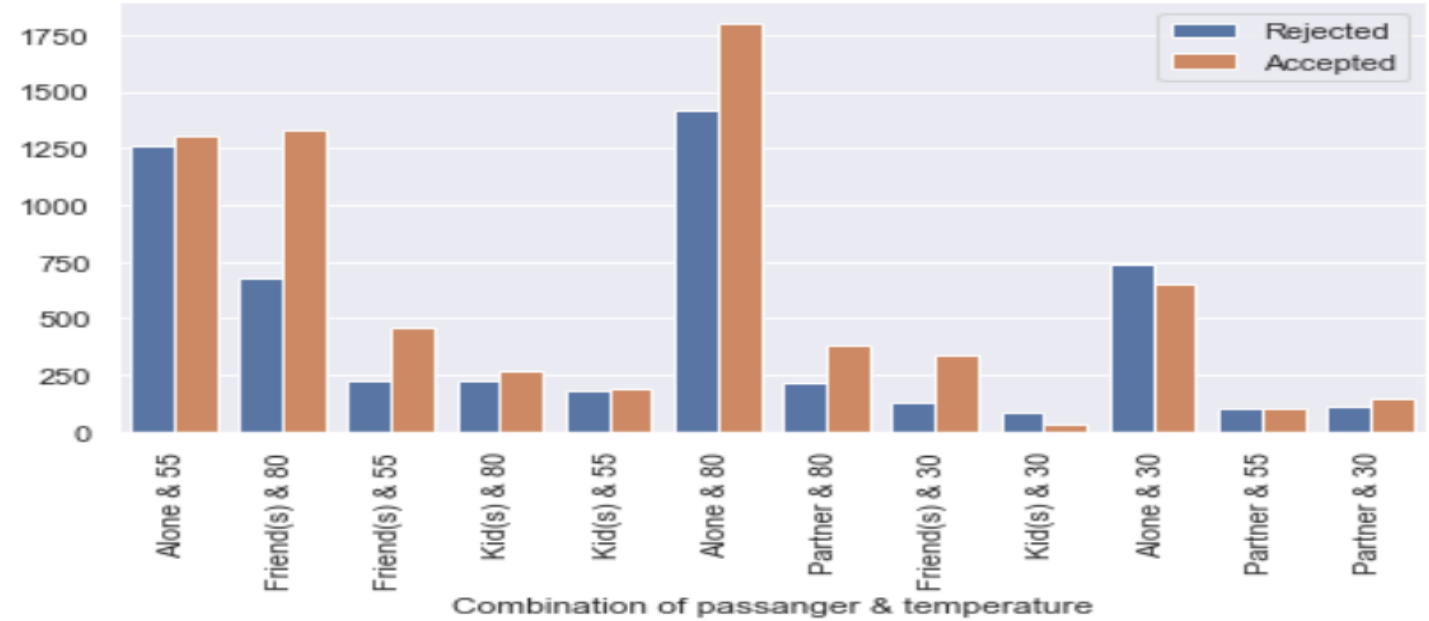
temperature

temperature wise accepted coupons

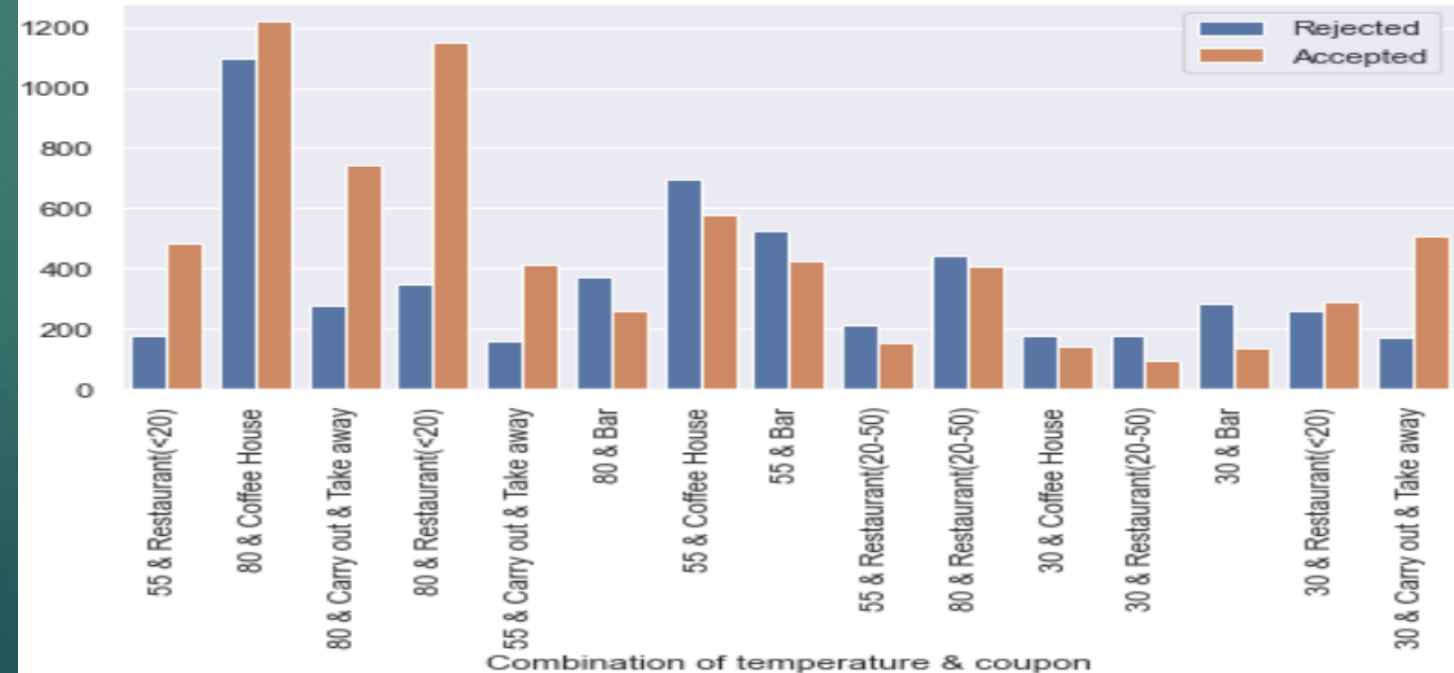


	temperature	Total_Count	% of Total	Accepted	Rejected	%Accepted	%Rejected
0	High	6328	51.061083	3786	2542	59.829	40.171
2	Medium	3828	18.050512	2055	1773	53.683	46.317
1	Low	2237	30.888405	1171	1066	52.347	47.653

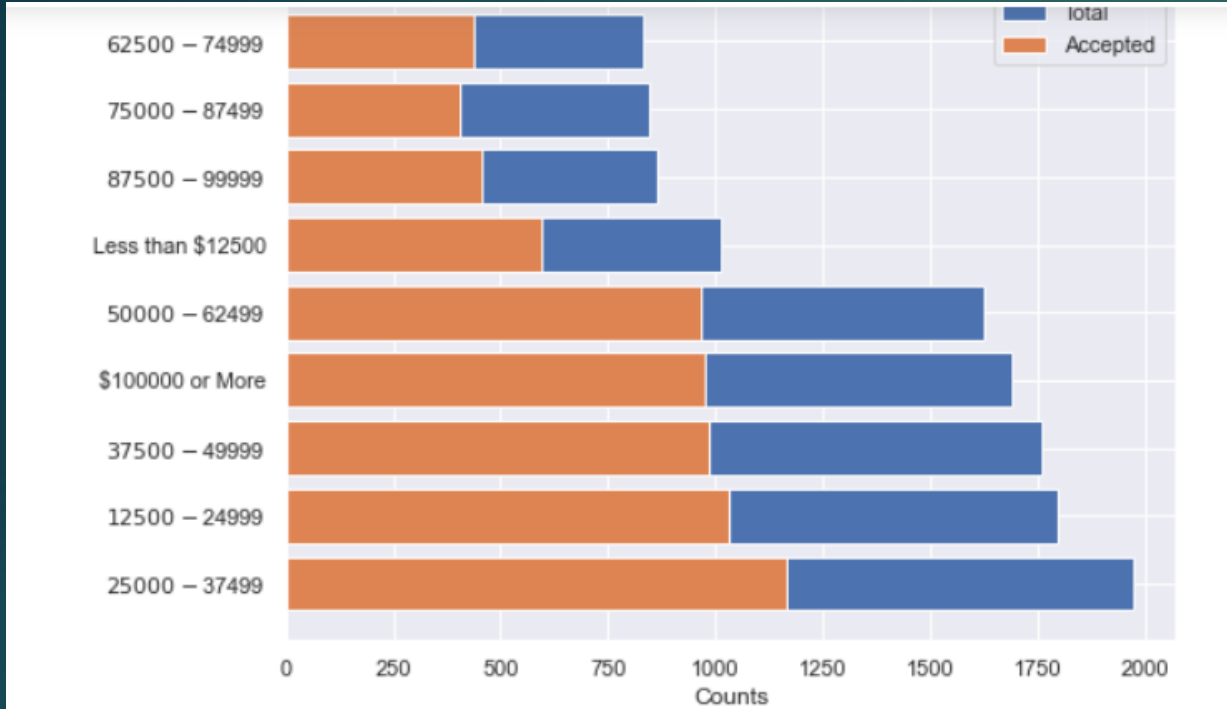
Effect of passanger & temperature on target variable



Effect of temperature & coupon on target variable



Income and occupation_class



1]:

	income	Total_Count	% of Total	Accepted	Rejected	%Accepted	%Rejected
2	25000-37499	1972	14.201565	1165	807	59.077	40.923
1	12500-24999	1795	14.483983	1029	766	57.326	42.674
3	37500-49999	1760	13.620592	987	773	56.080	43.920
0	\$100000 or More	1688	15.912209	974	714	57.701	42.299
4	50000-62499	1624	13.104172	964	660	59.360	40.640
8	Less than \$12500	1013	6.713467	593	420	58.539	41.461
7	87500-99999	865	6.810296	456	409	52.717	47.283
6	75000-87499	844	6.979747	407	437	48.223	51.777
5	62500-74999	832	8.173969	437	395	52.524	47.476



	occupation_class	Total_Count	% of Total	Accepted	Rejected	%Accepted	%Rejected
4	Medium_Low_Acceptance	3236	7.810861	1780	1456	55.006	44.994
2	Medium_Acceptance	3163	20.535786	1814	1349	57.351	42.649
3	Medium_High_Acceptance	2545	20.019366	1535	1010	60.314	39.686
1	Low_Acceptance	2481	25.522472	1236	1245	49.819	50.181
0	High_Acceptance	968	26.111515	647	321	66.839	33.161

Conclusion

- As we seen from previous slide the feature importance from random XGBClassifier, now we make conclusions
 1. coupon_freq has highest weightage in making coupon acceptance/rejection prediction, people with coupon_freq of 4-8, gt8, 1-3 accept coupons more frequently, people with less1 frequency accepts 50% times and people with freq never accepts only 22%, so we need to focus on these people, so that they can visit more frequently and motivate them for coupon acceptance.
 2. When validity of coupon 1d people accept more than 2h. So we should focus on coupons with validity of 1d.
 3. Carry out and Take away, Restaurant(<20) have higher acceptance ratio, so, focus should be on them for further business growth. Similarly Bar and Restaurant(20-50) were have least acceptance ratio, so need to focus on reason behind less productivity.
 4. When weather is Sunny people accepts more coupon than snowy and Rainy weather. Although in Sunny weather coupon acceptance rate is good but, when distance is more than 25minutes, this acceptance below 50%. So we need to motivate people to accept coupon in Snowy and Rainy weather and people having driving distance more 25 minutes in Sunny weather.
 5. We can see that coupon acceptance decreases with increase in driving distance, so we need to focus on motivating people with more driving distance to accept more coupon.
 6. In marital status, Single, unmarried partner, married partner and Divorced have more than 50% acceptance rate, while widowed people have less than 50% acceptance rate and widowed males acceptance ratio is less than 50% while in general males acceptance ratio is 58%.