Will the customer accept the coupon?

ML Classification Project

Paweł Krępic
2025

Project *agenda*

1. Research Problem & Goals
2. Dataset Description
3. Data Preprocessing & Exploratory Analysis
4. Model Evaluation and Comparison
5. Conclusions & Business Recommendations

Coupons are a common strategic marketing tool to influence consumer behavior. They help brands attract new customers, build loyalty with existing ones, and ultimately boost sales and revenue. They also provide valuable insights into customer preferences.

The challenge is to design **an effective coupon recommendation strategy**. Customers have different preferences and respond differently to promotional offers.

It's crucial to offer the **right coupon** to the **right person** – to increase redemption rates and encourage repeat usage.

Technical Goal

Build ML classification models to predict coupon acceptance based on different driving scenarios

Business Goal

Improve coupon redemption rates by predicting which customers are most likely to accept coupons

DATASET

Description

- The dataset comes from the **UC Irvine Machine Learning Repository** and was collected via survey on **Amazon Mechanical Turk**

- The survey examines different driving scenarios (including the destination, current time, weather conditions, passenger information, etc) and then asks drivers whether they would accept a recommended coupon

- The dataset contains **12 684 observations** and **25 features** (presented on the next slides) and the **target variable Y**.

Data is available here:
https://archive.ics.uci.edu/dataset/603/in+vehicle+coupon+recommendation

Feature	Description
destination	Home / Work / No Urgent Place
passenger	Alone / Friend(s) / Kid(s) / Partner *(feature meaning: who are the passengers in the car)*
weather	Sunny / Rainy / Snowy
temperature	30 / 55 / 80 *(degrees Fahrenheit)*
time	7AM / 10AM / 2PM / 6PM / 10PM
coupon	Bar / Coffee House / Carry out & Take away / Restaurant (<$20) Restaurant($20-$50) *(feature meaning: recommended coupon type)*
expiration	2h / 1d *(feature meaning: does the coupon expire in 1 day or in 2 hours?)*
gender	Female / Male
age	below21 / 21 / 26 / 31 / 36 / 41 / 46 / 50plus
maritalStatus	Single / Unmarried partner / Married partner / Divorced / Widowed
has_children	0 (No) / 1 (Yes)
education	Some High School / High School Graduate / Some college - no degree / Associates degree / Bachelors degree / Graduate degree (Masters or Doctorate)
car	Scooter and motorcycle / Mazda5 / do not drive / crossover / Car that is too old to install Onstar :D *(feature meaning: type of vehicle driven)*

Feature	Description
occupation	**25 unique values:** Unemployed / Architecture & Engineering / Student / Education&Training&Library / Healthcare Support / Healthcare Practitioners & Technical / Sales & Related / Management / Arts Design Entertainment Sports & Media / Computer & Mathematical / Life Physical Social Science / Personal Care & Service / Community & Social Services / Office & Administrative Support / Construction & Extraction / Legal / Retired / Installation Maintenance & Repair / Transportation & Material Moving / Business & Financial / Protective Service / Food Preparation & Serving Related / Production Occupations / Building & Grounds Cleaning & Maintenance / Farming Fishing & Forestry
income	Less than $12500 / $12500 - $24999 / $25000 - $37499 / $37500 - $49999 / $50000 - $62499 / $62500 - $74999 / $75000 - $87499 / $87500 - $99999 / $100000 or More)
Bar	never, less1, 1-3, 4-8, gt8 *(feature meaning: how many times does the driver go to a bar every month?)*
CoffeeHouse	never, less1, 1-3, 4-8, gt8 *(feature meaning: how many times does the driver go to a coffeehouse every month?)*
CarryAway	never, less1, 1-3, 4-8, gt8 (feature meaning: how many times does the driver get take-away food every month?)
RestaurantLessThan20	never, less1, 1-3, 4-8, gt8 *(feature meaning: how many times does the driver go to a restaurant with an average expense per person of less than $20 every month?)*
Restaurant20to50	never, less1, 1-3, 4-8, gt8 *(feature meaning: how many times does the driver go to a restaurant with average expense per person of $20 - $50 every month?)*
toCoupon_GEQ5min	0 (No) / 1 (Yes) *(feature meaning: driving distance to the restaurant/bar for using the coupon is greater than 5 minutes)*
toCoupon_GEQ15min	0 (No) / 1 (Yes) *(feature meaning: driving distance to the restaurant/bar for using the coupon is greater than 15 minutes)*
toCoupon_GEQ25min	0 (No) / 1 (Yes) *(feature meaning: driving distance to the restaurant/bar for using the coupon is greater than 25 minutes)*
direction_same	0 (No) / 1 (Yes) *(feature meaning: whether the restaurant/bar is in the same direction as the driver's destination)*
direction_opp	0 (No) / 1 (Yes) *(feature meaning: whether the restaurant/bar is in the opposite direction to the driver's destination)*

DATA PREPROCESSING

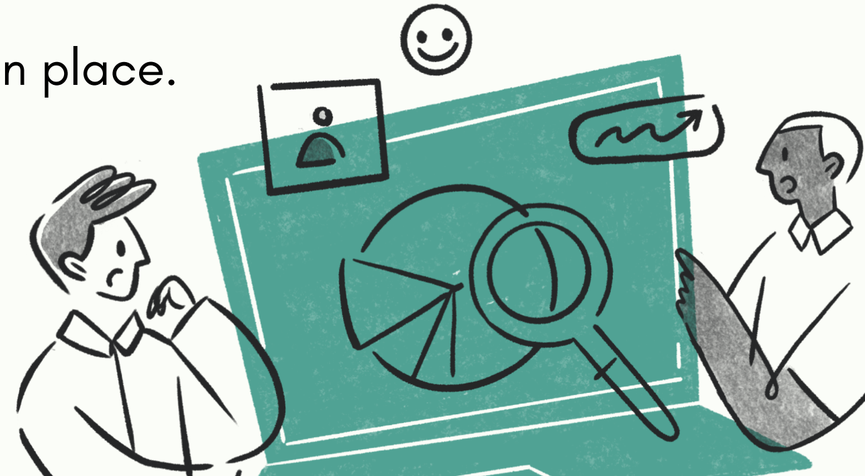
Key data transformations (part 1)

- **Removed irrelevant or redundant features** (*Car*, *toCoupon_GEQ5min*, *direction_opp*).

- **Dropped rows with missing values** (<5% of all observations).

- **Created new aggregated features:**
  a) *occupation_grouped* - broader job categories.
  b) *toCoupon_distance* - combined distance-related variables.
  c) visit_freq - summarized frequency of visits to the coupon redemption place.

- **Simplified some categorical variables** (*income*, *age*, *marital status*, *education*) by merging similar categories.



EDA

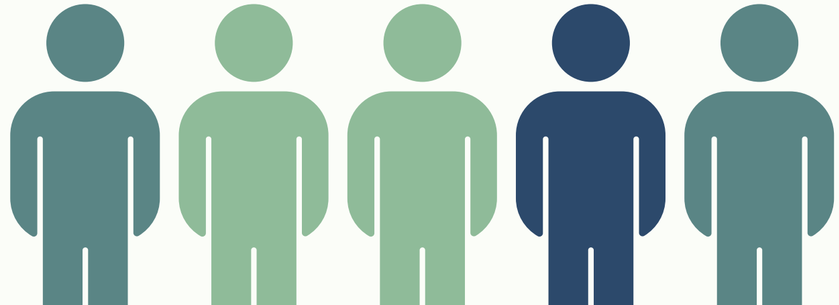
Some first patterns identified in coupon acceptance behaviour

Trip context: Coupons are more often accepted without a specific trip destination and less often when drivers are going Home or to Work. Traveling with friends is linked to higher acceptance, while traveling with kids lowers it. Time also matters – acceptance peaks in the afternoon, and is lower early in the morning and late in the evening.

External conditions: Sunny weather, higher temperature (80°F), and longer expiration time (1 day) are associated with higher acceptance.

Coupon type & frequency: Shorter distances and more frequent visits strongly increase acceptance rates. The most popular coupons are Carry out & Take away and Restaurant(<20).

Customer profile: Younger, single drivers without children accept coupons more often. Higher acceptance is also observed among students and physical workers, while retired drivers are least likely to accept coupons.



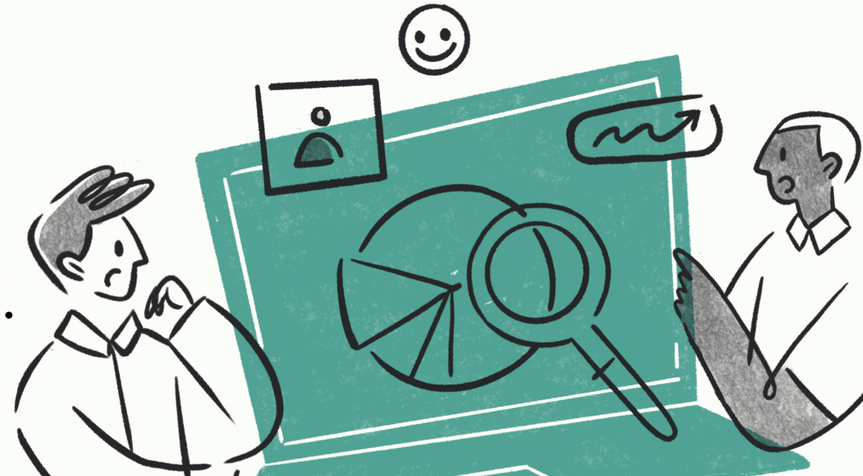
DATA PREPROCESSING

Key data transformations (part 2)

- Based on the initial exploratory analysis, **additional transformations** were applied to the variables *weather*, *temperature*, *and visit_freq*.

- **Correlation analysis (Cramer's V)** revealed strong redundancy, which led to **the removal of** the *destination* feature due to its overlap with the features *time* and *passenger*.

- After all transformations, the obtained set is ready for modeling and consists of **12 079 observations** and includes **17 features** (all treated as nominal) + **target Y**.



FEATURES for MODELING

After preprocessing and exploratory analysis

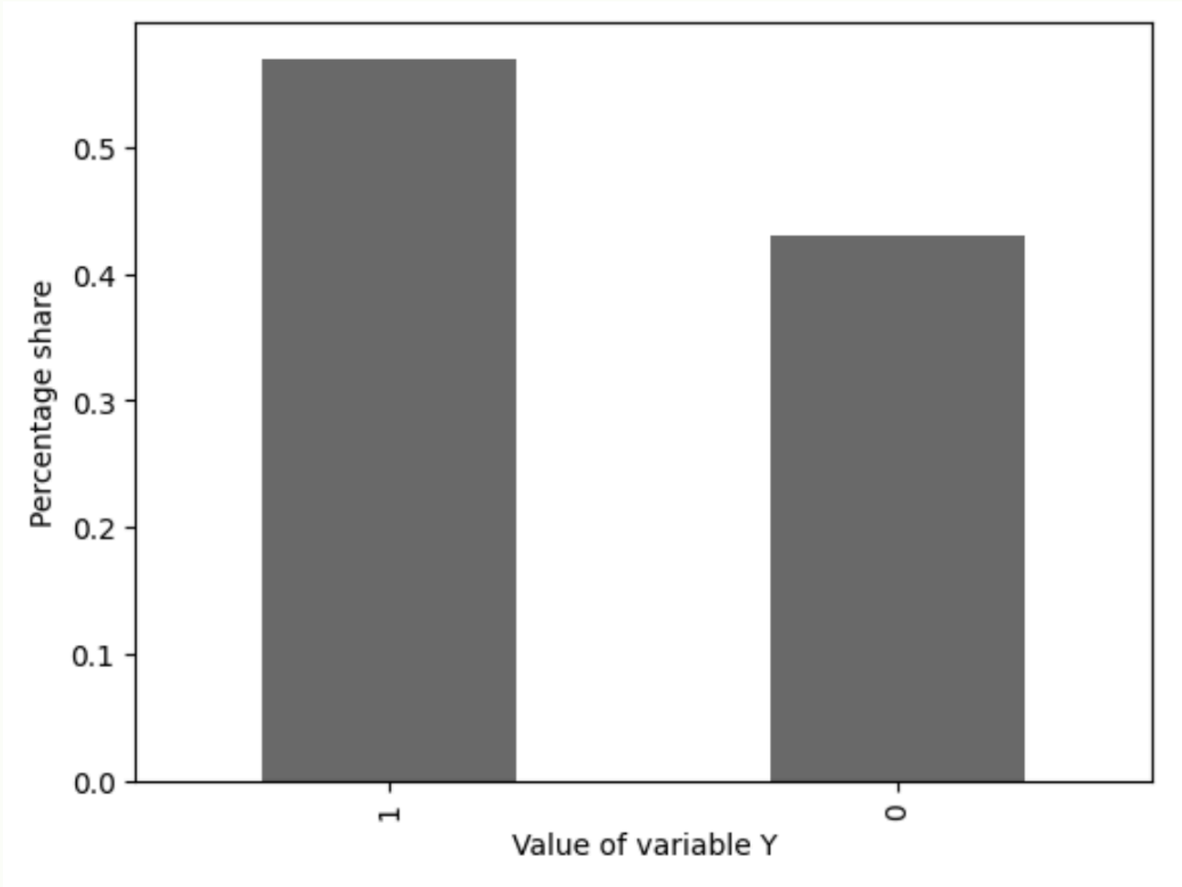
- **passanger:** Alone / Friend(s) / Kid(s) / Partner
- **weather:** Bad / Good
- **time:** 7AM / 10AM / 2PM / 6PM / 10PM
- **coupon:** Bar / Coffee House / Carry out & Take away / Restaurant (<$20) / Restaurant($20-$50)
- **expiration:** 2h / 1d
- **gender:** Female / Male
- **age:** >21, 21-30, 31-40, 41-50, 50+
- **maritalStatus:** Single / Has_Partner / Had_Partner
- **has_children:** 0 (No) / 1 (Yes)
- **education:** Low / Mid / High
- **income:** 1 (lowest group) / 2 (medium group) / 3 (highest group)
- **direction_same:** 0 (No) / 1 (Yes)
- **occupation_grouped:** Student / Unemployed / Physical / Business_Services / Social_Edu / Tech_Eng / Retired
- **toCoupon_distance:** 1 (shortest) / 2 (medium) / 3 (longest)
- **visit_freq:** 0 (never) / 1 (less than 1) / 2 (1-3) / 3 (4 or more)
- **temp_outside:** Cold / Warm

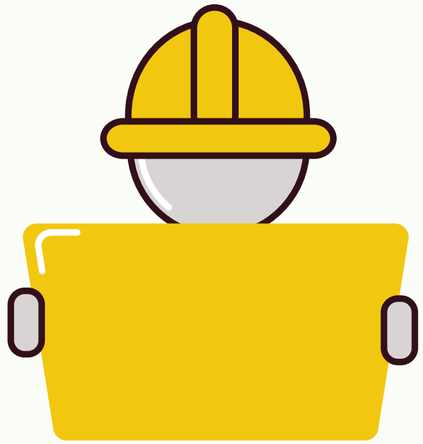
TARGET VARIABLE Y

Y - Was the coupon accepted?

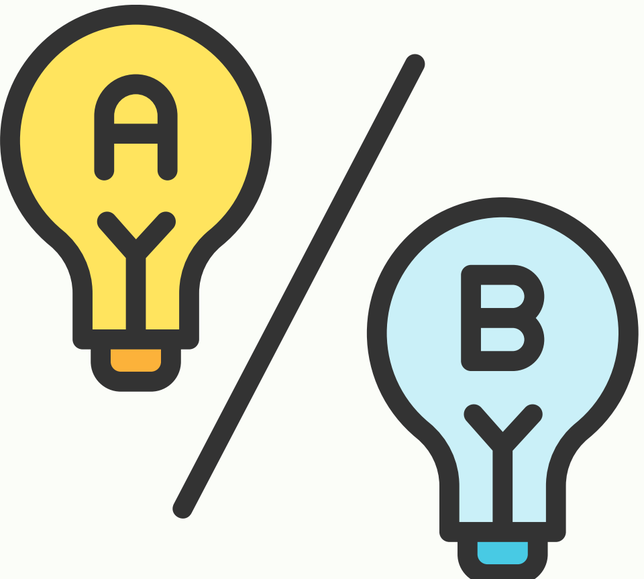
- **1** – coupon accepted (**56,93%**)
- **0** – coupon rejected (**43,07%**)

The sample is quite well balanced.





MODEL EVALUATION AND COMPARISON

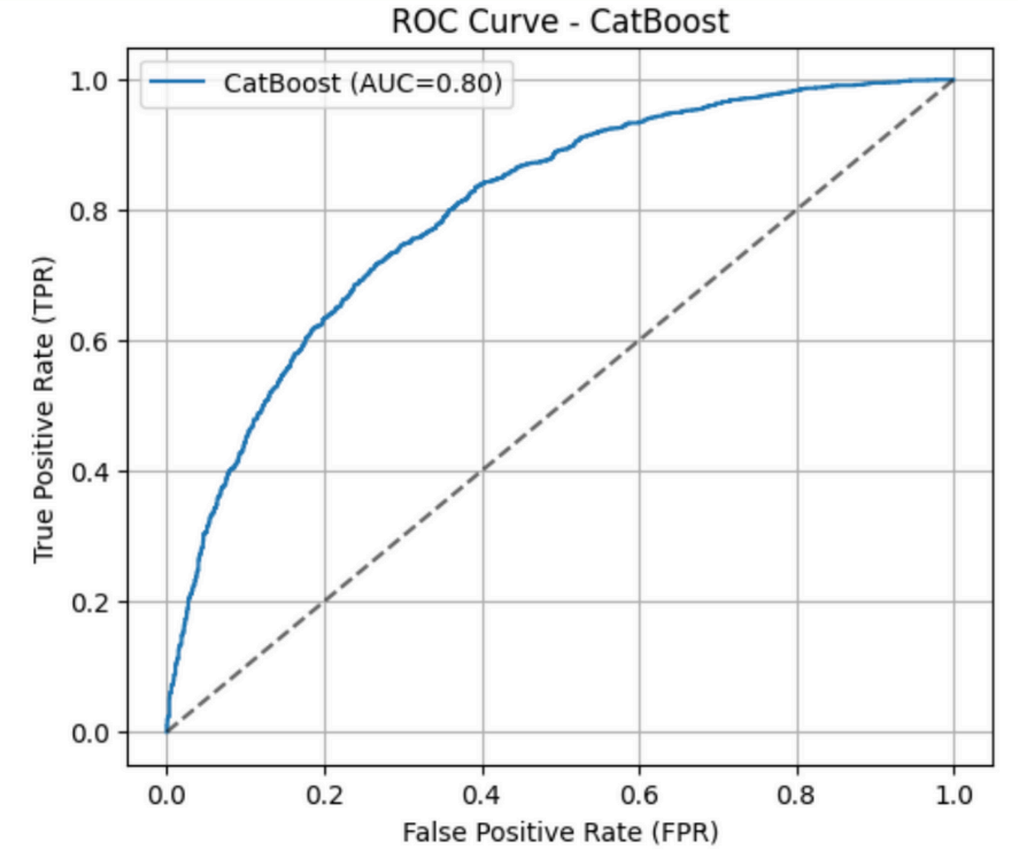


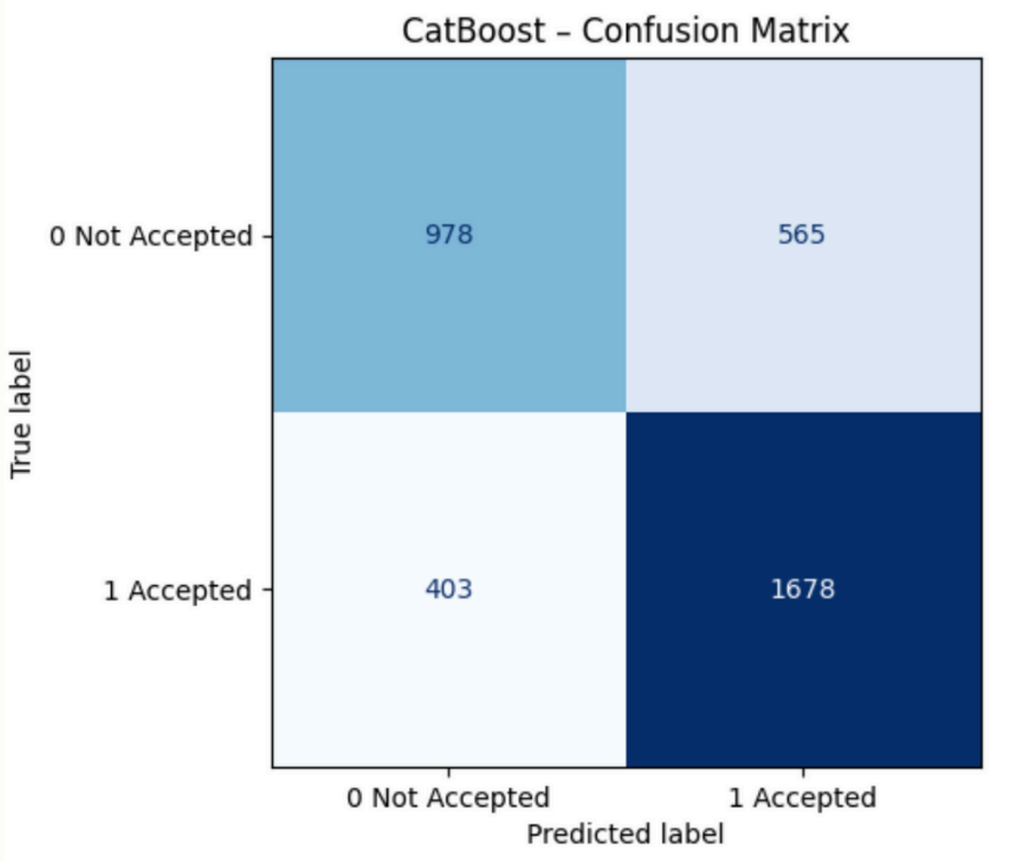
MODELS *and their results*

	Logistic Regression	Decision Tree	Random Forest	Gradient Boosting	XGBoost	CatBoost
	5	6	4	3	2	1
F1-Score	0.76	0.76	0.77	0.78	0.77	0.78
Accuracy	0.71	0.69	0.72	0.72	0.73	0.73
ROC-AUC	0.76	0.74	0.78	0.78	0.79	0.80
Evaluation	☆	☹	☆ ☆	☆ ☆ ☆	☆ ☆ ☆ ☆	☆ ☆ ☆ ☆ ☆

CatBoost

ROC Curve and Confusion Matrix

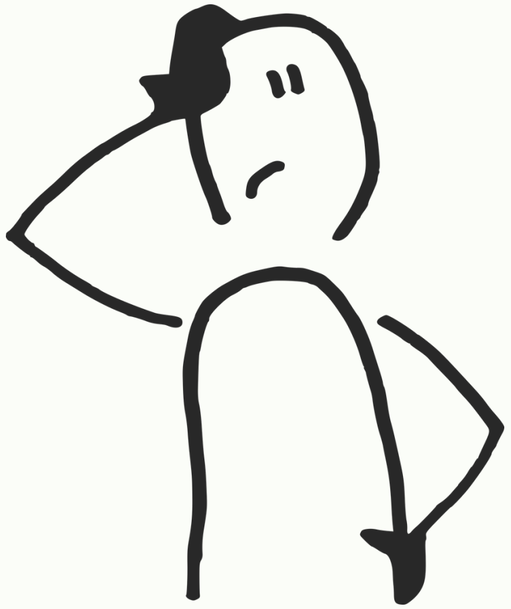




CATBOOST

Feature Importance (TOP 15)

	Feature	Importance
1.	visit_freq_2	9.82
2.	visit_freq_2	8.85
3.	expiration_2h	5.95
4.	visit_freq_1	5.10
5.	coupon_Restaurant(<20)	4.38
6.	gender_Male	3.88
7.	coupon_Carry out & Take away	3.69
8.	coupon_Coffee House	3.24
9.	income_2	3.22
10.	education_Mid	3.05
11.	age_31-40	3.02
12.	has_children_1	2.94
13.	passanger_Friend(s)	2.83
14.	time_7AM	2.44
15.	temp_outside_Warm	2.29



Catboost feature importance values represent the relative contribution of each feature to the predictive performance of the model. They indicate which effects are the strongest.

However, these values are **difficult to interpret** and **do not indicate the direction** of a given effect (whether a feature increases or decreases the chances of coupon acceptance). For assessing **the strength** and **direction** of the impact of individual features, the **logistic regression results are useful**.

Logistic Regression

Feature Importance (TOP 25)

	Feature	Reference	Coefficient (Log odds)	Exp(Coefficient) (Odds)	% increase / descrease in odds
1.	visit_freq_3	visit_freq_0	2.09	8.08	708 %
2.	visit_freq_2	visit_freq_0	1.93	6.89	589 %
3.	visit_freq_1	visit_freq_0	1.36	3.90	290 %
4.	expiration_2h	expiration_1d	-0.89	0.41	59 %
5.	coupon_Carry out & Take away	coupon_Bar	0.74	2.10	110 %
6.	time_10PM	time_10AM	-0.67	0.51	49 %
7.	passanger_Friend(s)	passanger_Alone	0.67	1.95	95 %
8.	coupon_Restaurant(<20)	coupon_Bar	0.59	1.80	80 %
9.	passanger_Partner	passanger_Alone	0.51	1.67	67 %
10.	time_7AM	time_10AM	-0.50	0.61	39 %
11.	weather_Good	weather_Bad	0.47	1.60	60 %
12.	education_Low	education_High	0.37	1.48	48 %
13.	income_2	income_1	-0.37	0.69	31 %

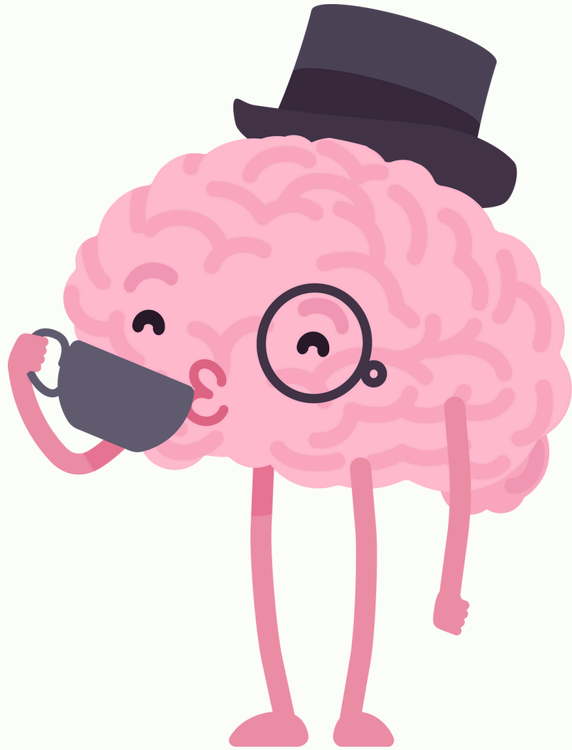
LOGISTIC REGRESSION

Feature Importance (TOP 25)

	Feature	Reference	Coefficient (Log odds)	Exp(Coefficient) (Odds)	% increase / descrease in odds
14.	maritalStatus_Single	maritalStatus_Had_ Partner	0.36	1.43	43 %
15.	time_2PM	time_10AM	-0.31	0.73	27 %
16.	direction_same_1	direction_same_0	0.31	1.36	36 %
17.	age_>21	age_21-30	0.24	1.27	27 %
18.	time_6PM	time_10AM	-0.24	0.79	21%
19.	occupation_grouped_ Tech_Eng	occupation_Business_Services	0.20	1.22	22 %
20.	occupation_grouped_ Retired	occupation_Business_Services	-0.20	0.82	18 %
21.	toCoupon_distance_3	toCoupon_distance_1	-0.18	0.84	16 %
22.	age_31-40	age_21-30	-0.18	0.84	16 %
23.	occupation_grouped_ Physical	occupation_Business_Services	0.16	1.17	17 %
24.	gender_Male	gender_Female	0.16	1.17	17 %
25.	age_50+	age_21-30	-0.15	0.86	14 %

conclusions

and Business Recommendations



The **CatBoost** model achieved the **best performance** (F1-Score, Accuracy, ROC-AUC), with all models reaching at least 0.7 on key metrics (except Accuracy for Decision Tree).

Logistic Regression coefficients were used to **interpret the direction and strength** of individual factors.

COUPON ACCEPTANCE

Examples of positive factors

Physical and technical & engineering jobs

Destination aligned with trip direction

Being single

Younger age (under 30)

Frequent visits to the coupon location

Longer coupon validity

Traveling with friends or a partner

Sunny weather

Lower od Mid education levels

Being Male

Carry out & Take away coupon type

COUPON ACCEPTANCE

Examples of negative factors

No visiting the coupon location

Early morning or late evening hours

Destination not aligned with trip direction

Being Retired

Being older (50+)

Shorter coupon validity

Being Female

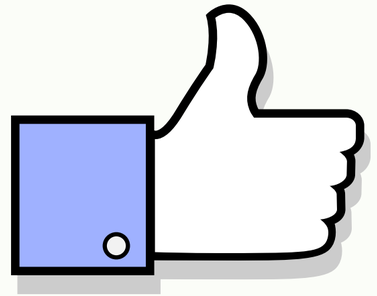
Rainy or snowy day

High education level

Longer distance to the coupon location

Conclusions

and Business Recommendations



- **Target the right audience** → Younger drivers, especially those traveling with friends or a partner and frequently visiting the coupon redemption place, are the most responsive.

- **Offer what works best** → Carry out & Take away and Restaurant(<20) coupons show the highest acceptance rates.

- **Leverage timing & context** → Sunny days boost acceptance — *"Sunny afternoon? Perfect time for coffee!"* Avoid very early mornings and late evenings when drivers are focused on commuting.

- **Keep it close** → Coupons tied to nearby locations (like the nearest restaurant) are more likely to be accepted.

Next step: Explore alternative strategies for harder-to-reach groups — such as older drivers or those living far from coupon locations (maybe online delivery offers?).

THANK YOU

for watching!