

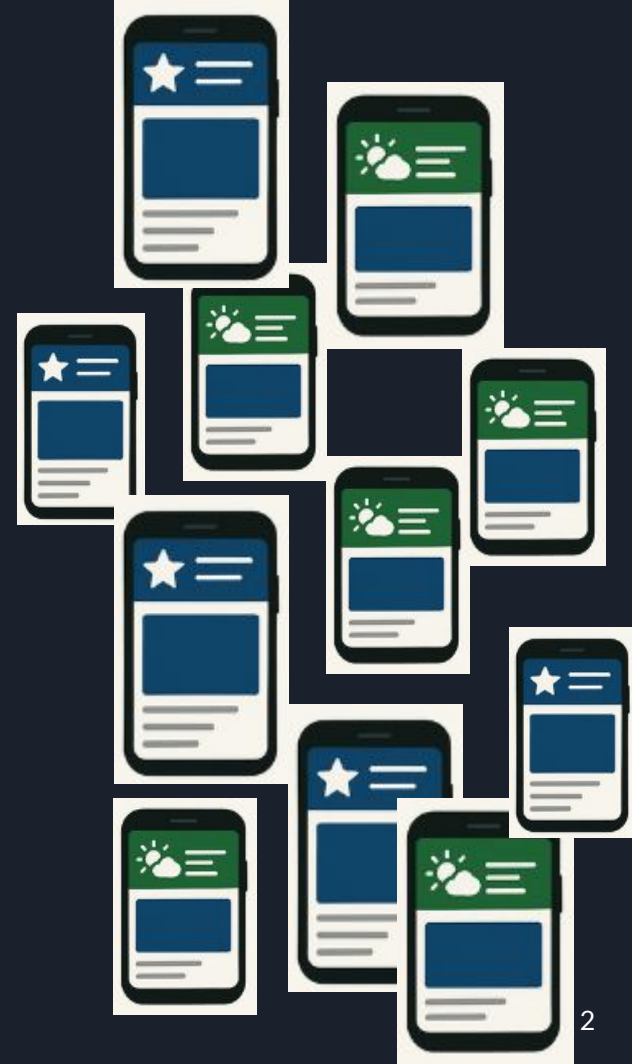
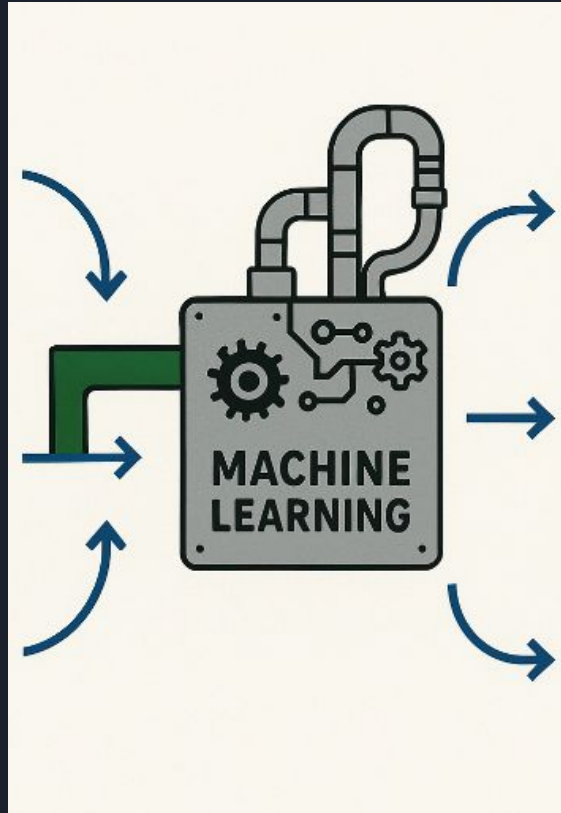
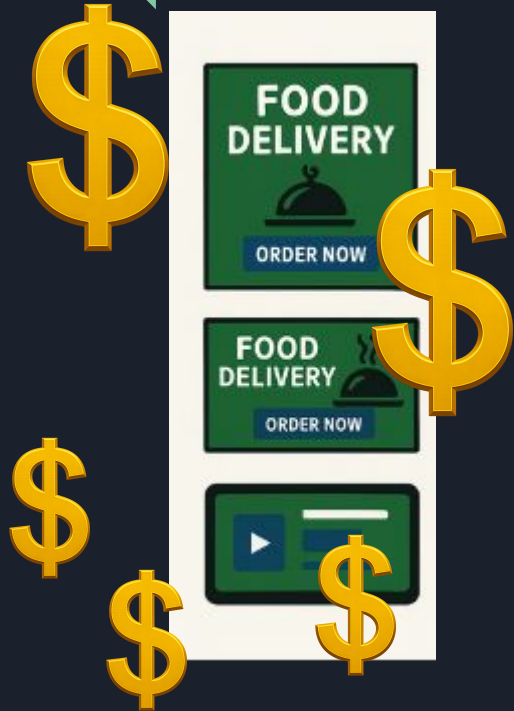


Predicting Purchases

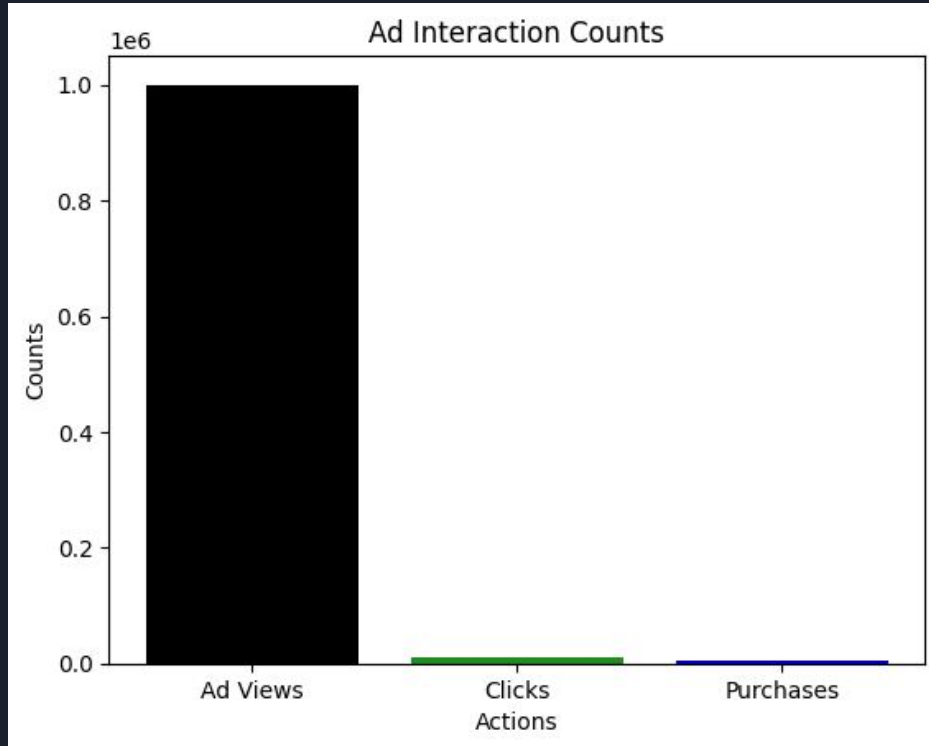
Machine Learning as a
Service

Debora Bimbi
18.04.2025

Machine Learning as a Service



User Interaction with Ads



So many categorical values

```
[  
  'en', 'zh', 'hi', 'es', 'fr',  
  'ar', 'bn', 'ru', 'pt', 'id',  
  'ur', 'de', 'ja', 'sw', 'mr',  
  'te', 'ta', 'tr', 'ko', 'it'  
]
```

```
[  
  "Food Delivery",  
  "Health & Fitness",  
  "Health & Medicine",  
  "Personal Finance",  
  "Dating",  
  "Gaming",  
  "News & Weather"  
]
```

Time of day

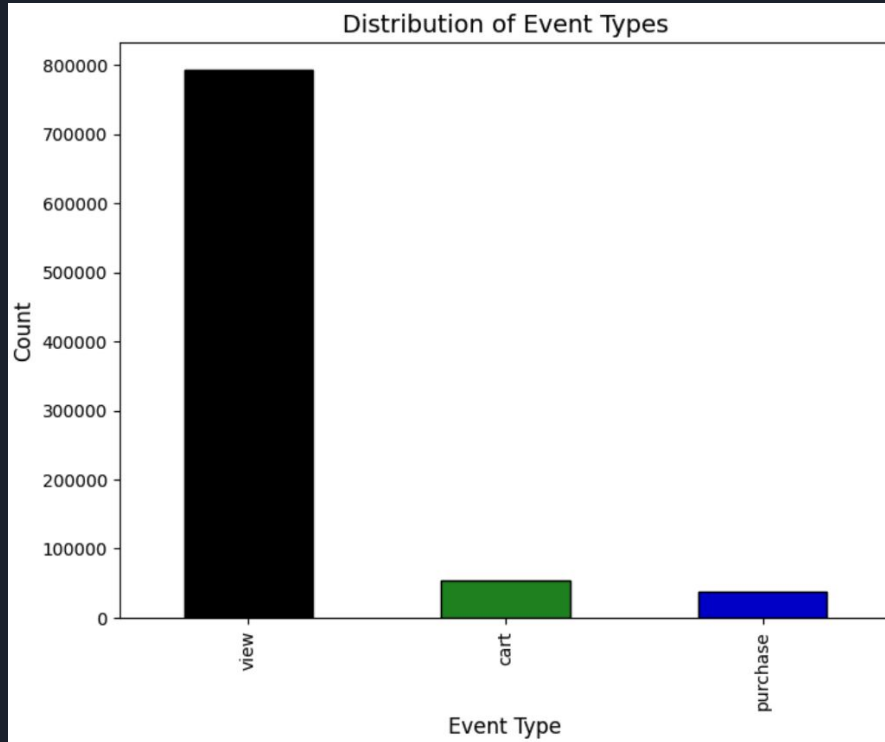
Day of the Week

```
[  
  'Samsung Galaxy S21',  
  'Samsung Galaxy Note 20',  
  'Google Pixel 6',  
  'OnePlus 9 Pro',  
  'Xiaomi Mi 11',  
  'Oppo Find X3 Pro',  
  'Huawei P40 Pro',  
  'Sony Xperia 5 II',  
  'Motorola Edge Plus',  
  'LG Velvet',  
  'iPhone 13',  
  'iPhone 13 Mini',  
  'iPhone 13 Pro',  
  'iPhone 13 Pro Max',  
  'iPhone 12',  
  'iPhone 12 Mini',  
  'iPhone 12 Pro',  
  'iPhone SE (2020)',  
  'iPhone 11',  
  'iPhone XR'  
]
```

Similar Dataset

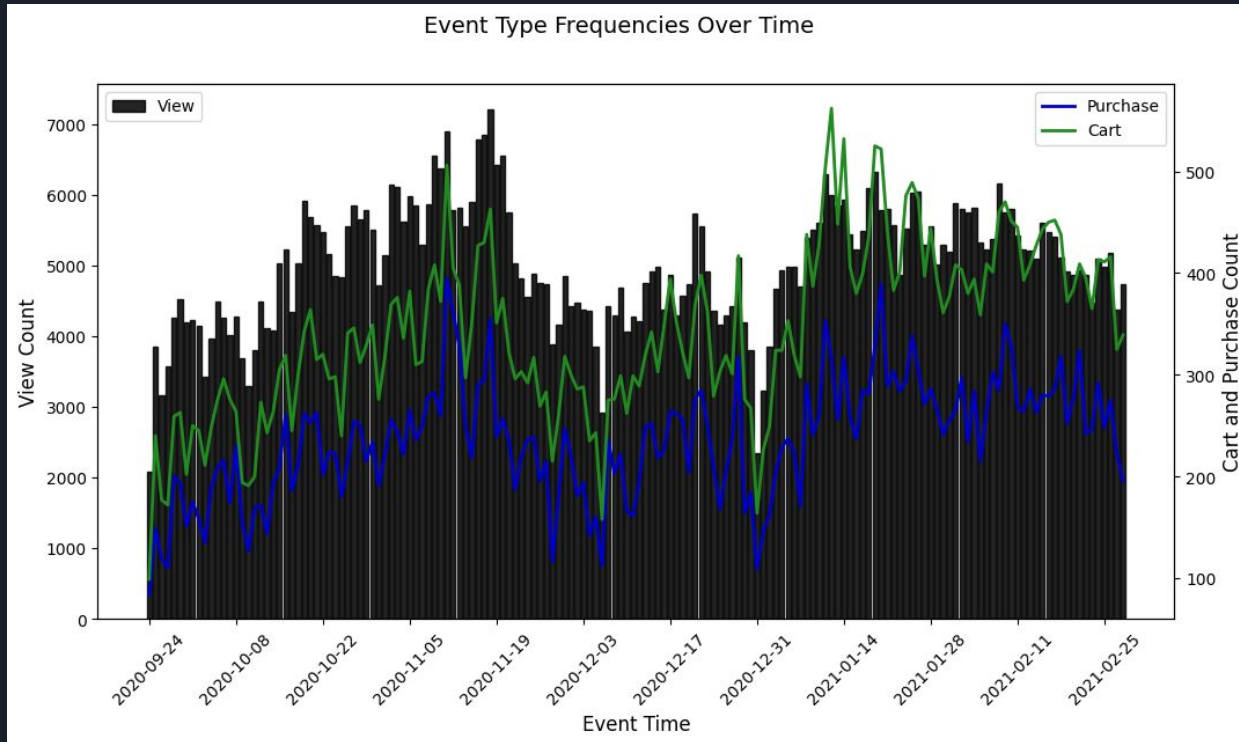
E-commerce:
Event_time
Event_type
Product_ID
Category_ID
Category_Code
Brand
Price
UserID

800 000 rows
From 2020-09-24
until 2021-02-28

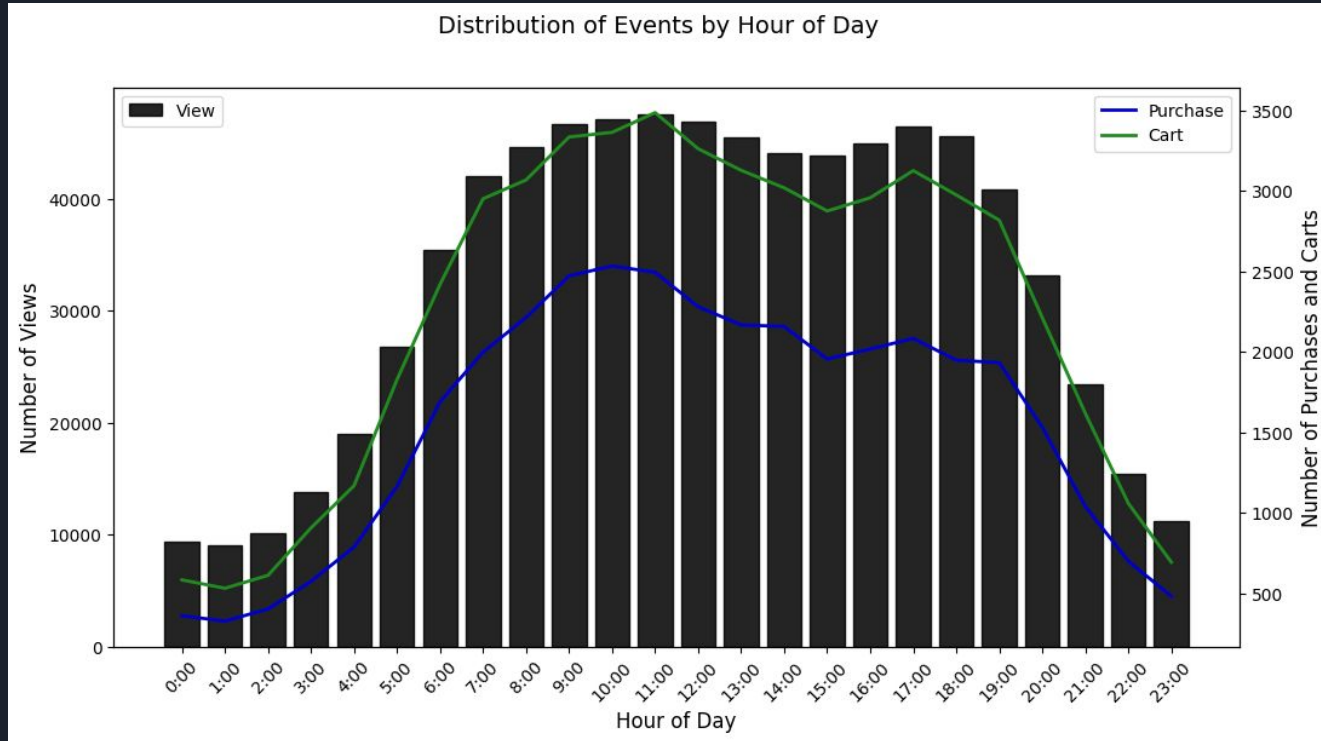


Predict
Views that
turn into
Purchases

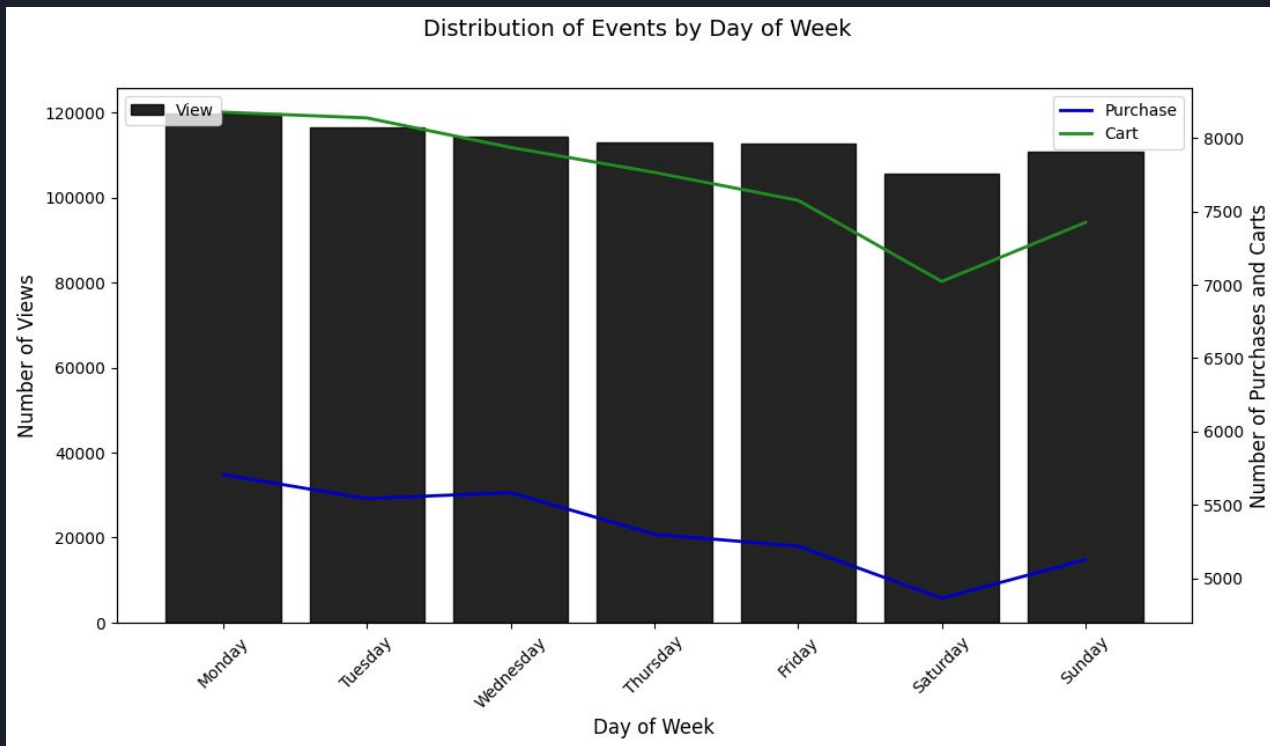
Time Series of Events



Feature Engineering: Time of Day (H)



Feature Engineering: Day of Week



Price range

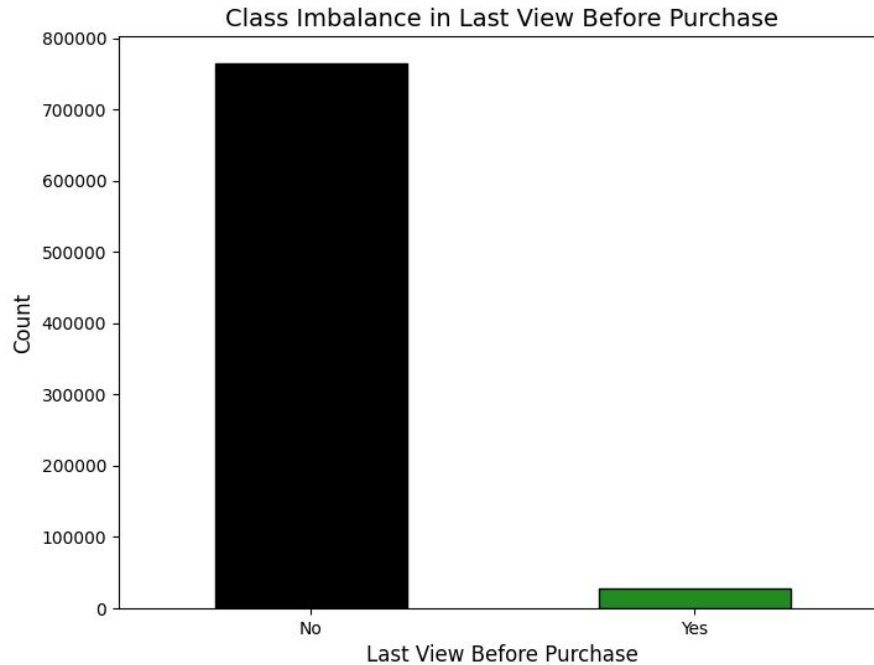
Views that don't lead to purchase



Views that lead to purchases



Final Class Imbalance





X \longrightarrow a lot of strings
y \longrightarrow 0 or 1

LabelEncoder()

Label Encodings:

```
product_id encoding: {0: '1000108', 1: '1000111', 2: '1000112', 3: '1000113', 4: '1000116', 5: '1000175',  
category_id encoding: {0: 2144415921001726020, 1: 2144415921085612102, 2: 2144415921169498184, 3: 2144415  
category_code encoding: {0: 'Unknown', 1: 'accessories.bag', 2: 'accessories.briefcase', 3: 'apparel.glov  
brand encoding: {0: 'Unknown', 1: 'a-data', 2: 'accesstyle', 3: 'accord', 4: 'acct', 5: 'acecamp', 6: 'ac  
last_view_before_cart encoding: {0: 0, 1: 1}  
already_in_cart encoding: {0: 0, 1: 1}  
time_of_day encoding: {0: '00:00:00', 1: '01:00:00', 2: '02:00:00', 3: '03:00:00', 4: '04:00:00', 5: '05:  
day_of_week encoding: {0: 'Friday', 1: 'Monday', 2: 'Saturday', 3: 'Sunday', 4: 'Thursday', 5: 'Tuesday',
```

StandardScaler()

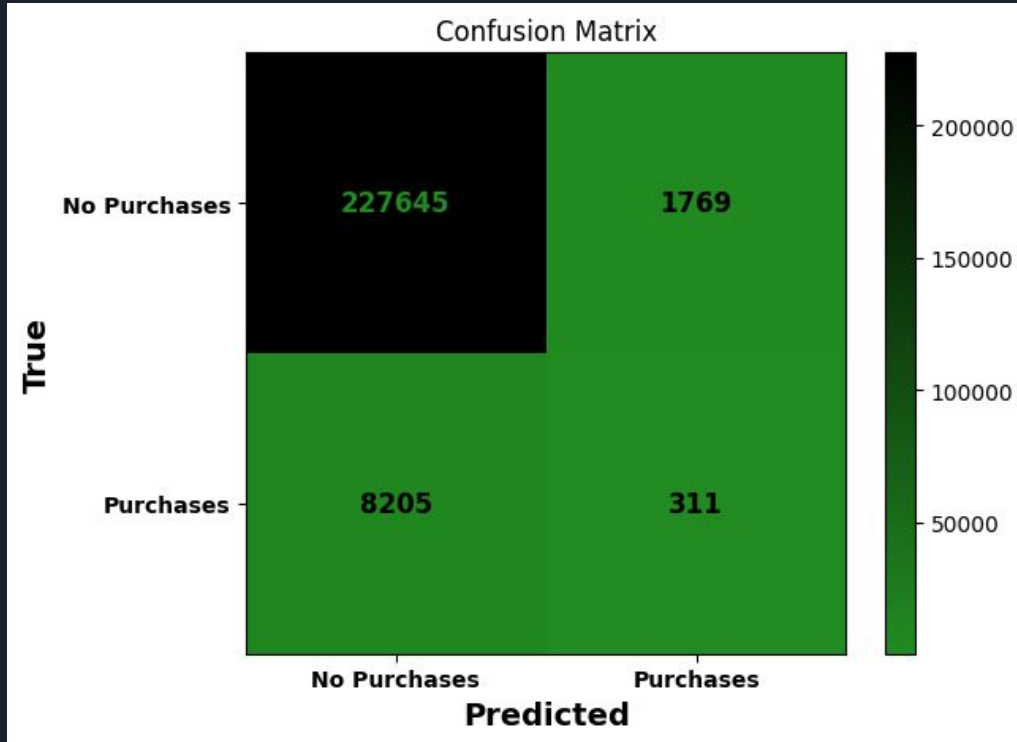


Best Model Results

PRECISION

Random Forest 0.15

Confusion Matrix

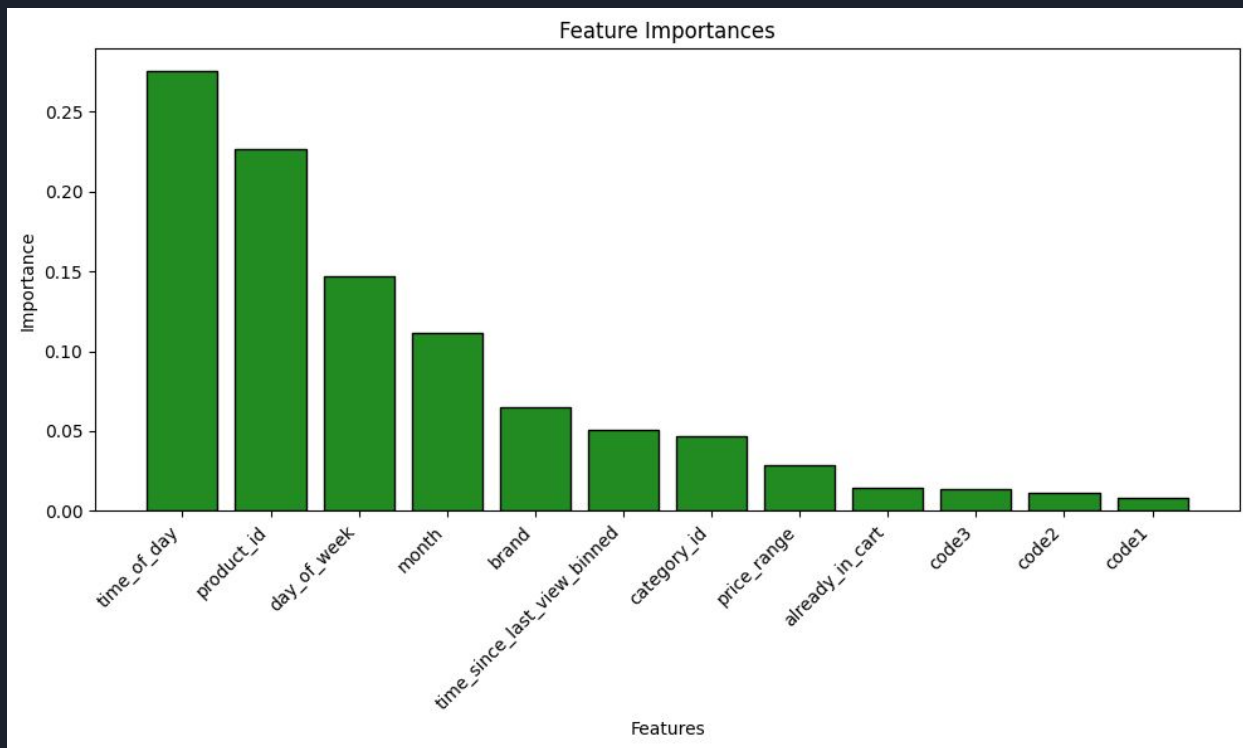


PRECISION

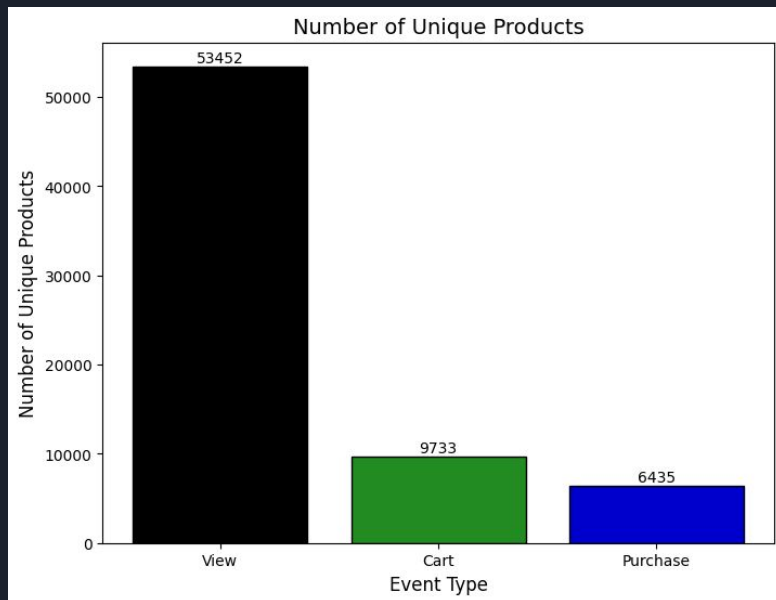
$$\frac{TP}{TP + FP}$$

$$\frac{311}{(311 + 1769)} = 0.1479$$

Feature Importance for RandomForestClassifier



Feature Engineering: Product_ID



THANK YOU