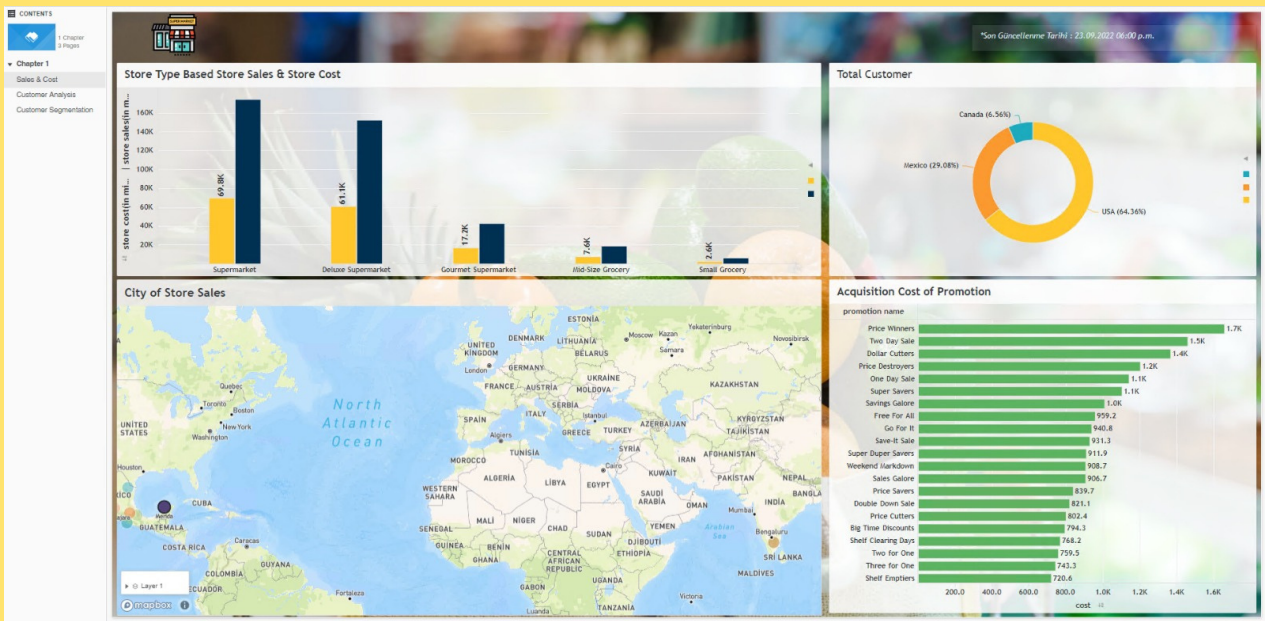


COST PREDICTION ON ACQUIRING CUSTOMERS

With 13.59 RMSE value, We estimate the cost that markets in every segment, from small to large, should allocate for media.

SALES AND COST ANALYSIS

- How is country-based market growth going, how is customer distribution?
- Sales according to store types and costs
- Easy and fast tracking is provided on the map
- Follow acquisition cost of promotions and budget them

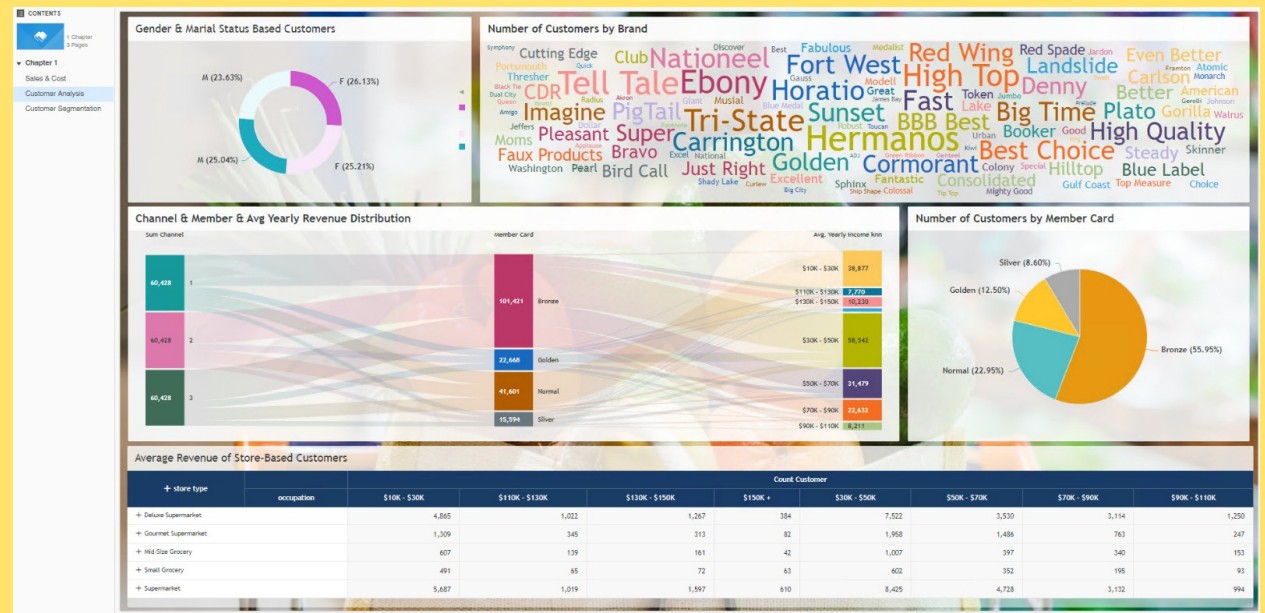


CUSTOMER ANALYSIS



Know your customers' preferences and develop sales strategies based on them

- Which brand are they more inclined to?
- Gender and marital status distribution
- What channels do they come from the most?
- Which membership card use more
- Yearly income distributions
- Average revenue coming from customers



SEGMENTATION AND PREDICTION ANALYSIS

- Customer segmentation by cost and marj
- How much of their budgets they allocate according to the store types they spend on the media?
- How much profit they make as a result of these expenditures?



If you want to take your place in the market and increase your profit, you can contact us!

Kübra Sak

Handan Yıldız

Reyhan Göger

Gizem Yılmaz

COST PREDICTION ON ACQUIRING CUSTOMERS

PREDICT COST ON MEDIA CAMPAIGNS IN FOOD MART OF USA .
ON THE BASIS OF **60K CUSTOMERS** INCOME ,PRODUCT,PROMOTION AND STORE
FEATURES.

STORY OF THE DATASET

ABOUT FOODMART:

Food Mart (CFM) is a chain of convenience stores in the United States. The private company's headquarters are located in Mentor, Ohio, and there are currently approximately 20 stores located in the US. Convenient Food Mart operates on the franchise system.

40 COLUMNS

60458 ROWS

16.67 MB SIZE

food_category	ürün kategorisi
food_department	ürünün satıldığı departman
food_family	ürün ailesi (Drink,Food,Nonconsumable)
store_sales(in millions)	toplam mağaza satışı (SRP*unit_sales(in millions))
store_cost(in millions)	toplam mağaza maliyeti
unit_sales(in millions)	satılan ürün adeti
promotion_name	promosyon ismi
sales_country	satış yapılan mağazanın ülkesi
marital_status	müşteri evlilik durumu
gender	müşteri cinsiyeti
total_children	müşterinin sahip olduğu çocuk sayısı
education	müşterinin eğitim durumu
member_card	müşterinin üyelik bilgisi
occupation	müşterinin mesleği (Manual,Skilled Manual,Professional,Management,Clerical)
houseowner	müşterinin ev sahibi olup olmadığı bilgisi
avg_cars_at home(approx)	müşterinin evinde sahip olduğu araç sayısı
avg_yearly_income	müşterinin ortalama yıllık gelirinin aralığı
num_children_at_home	müşterinin evinde onunla birlikte yaşayan çocuk sayısı
avg_cars_at home(approx).1	dublke kolon
brand_name	ürünün markası
SRP	önerilen satış fiyatı
gross_weight	ürünün brüt ağırlığı
net_weight	ürünün net ağırlığı
recyclable_package	ürünün geri dönüştürülebilen ambalaj bilgisi
low_fat	ürünün düşük kalorili olup olmaması
units_per_case	kutudan çıkan adet
store_type	mağazanın tipi(Deluxe Supermarket, Supermarket, Gourmet Supermarket, Mid-Size Grocery, Small Grocery)
store_city	mağazın şehri
store_state	mağazanın ilçesi
store_sqft	mağazanın alanı
grocery_sqft	mağazanın manav bölümünün alanı
frozen_sqft	mağazanın dondurulmuş bölümün alanı
meat_sqft	mağazanın dondurulmuş bölümün alanı
coffee_bar	mağazada coffe bar bölümü var mı?
video_store	mağazada video store bölümü var mı?
salad_bar	mağazada salad bar bölümü var mı?
prepared_food	mağazada hazır gıda bölümü var mı?
florist	mağazada çiçekçi bölümü var mı?
media_type	tanıtımlar nereden yapılıyor, medya kanalları
cost	medya harcamaları (Hedef değişken)

CORRELATION ANALYSIS

- Grossy_sqft x Meat_sqft **High Negative Correlation**
- Store_sales x Store_cost **High Positive Correlation**
- Store_sales x SRP **High Positive Correlation**
- Gross_weight x Net_weight **High Positive Correlation**
- Salad_bar x Prepared_food x Coffee_bar x Video_store x Florist **Meidum Positive Correlation**



DATA CLEANING

- There is no missing values
- There is no outlier

COST PREDICTION ON ACQUIRING CUSTOMERS

FEATURE ENGINEERING

- 41 new feature generated

NEW_MARGIN	üründen elde edilen kar
NEW_UNIT_PRICE_AT_BULK_SALES	toplu satıştaki birim fiyat(kutu içerisinde birden fazla ürün var)
NEW_COST_SALES_RATIO	maliyetin kazanç'a oranı
NEW_GROSS_NET_RATIO	brütün net'e oranı
NEW_WELFARE_LEVEL	refah seviyesi
NEW_MEMBER_CARD_LEVEL	üye card bilgileri
NEW_EDUCATION_LEVEL	eğitim seviyesi
NEW_STORE_TYPE_LEVEL	store type değişkenin ordinal ifadesi
MEDIA TYPE	değişkeni içerisinde değerlerin split edilip ayrı ayrı kolonların üretilmesi
NEW_ONLINE_CHANNEL	online kanallardan reklam verildi mi?
NEW_PROMOTION_DAY	promosyon içerisinde DAY veya WEEKEND geçen promosyonlar(Süre kısıtını ifade ediyor)
NEW_PROMOTION_SAVER	promosyon içerisinde SAVERS veya SAVING geçen promosyonlar(Tassarruf içeren ifade ediyor)
NEW_GROCERY_RATIO	manav alanının tüm mağazaya oranı
NEW_MEAT_RATIO	et alanının tüm mağazaya oranı
NEW_EXTRA_AREAS	kaç tane ek alan var('video_store', 'salad_bar', 'prepared_food', 'florist','coffee_bar')
NEW_STORE_NUM_IN_STATE	bulunduğu eyalette kaç tane mağaza var
NEW_STORE_REVENUE_IN_STATE	bulunduğu eyalette toplam gelir(geçici değişken daha sonra silinecek)
NEW_RATIO_SALES_TO_STATE	bulunduğu eyalete göre gelir oranı
SUM_CHANNEL	promosyonun çıktığı toplam kanal

FEATURE INTERACTION

NEW_FOOD_NUM_IN_BRAND_AT_STORE	İlgili mağazada Marka'ya ait kaç ürün var
NEW_REVENUE_IN_BRAND_AT_STORE	İlgili Mağazada Marka'ya ait gelir (Geçici değişken olarak kullanılacak)
NEW_RATIO_REVENUE_TO_BRAND_AT_STORE	Bulunduğu mağazada ilgili markaya ait toplam gelirden ne kadarı
NEW_FOOD_NUM_IN_BRAND	Marka'ya ait kaç ürün var
NEW_REVENUE_IN_BRAND	Marka'ya ait gelir(Geçici değişken olarak kullanılacak)
NEW_BRAND_BASED_REVENUE_BY_STORE	Bulunduğu mağazada ilgili markaya ait toplam gelirden ne kadarı
NEW_MARRIED_CHILDREN_RELATION	Çocuk evli olma durumu (Evli çocuklu, evli çocuksuz, bekar çocuklu, bekar çocuksuz)
NEW_RATIO_CHILDREN_AT_HOME	Toplam çocuğun evdeki çocuğa oranı

- One hot encoding

food_family
sales_country
marital_status
gender
occupation
houseowner
new_married_children_relation

STANDART SCALER

- Standardized the values
- StandardScaler method is used.

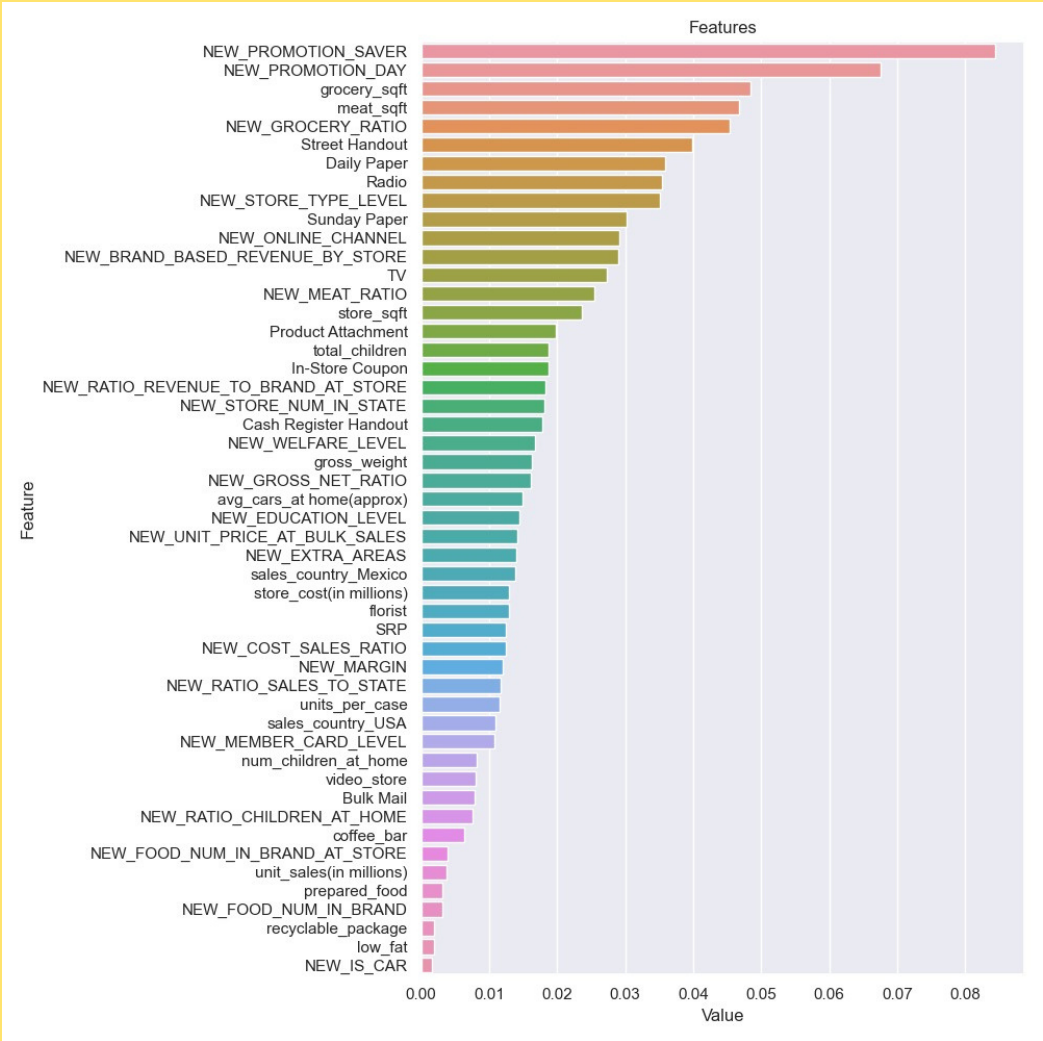
BASE MODELS

- Tried models are listed below. Random Forest gives the best result.

LR	28.2018
Lasso	29.3407
Ridge	28.3192
KNN	27.0757
CART	18.1334
RF	14.7013
LightGBM	16.8066
XGBoost	14.6911

COST PREDICTION ON ACQUIRING CUSTOMERS

FEATURE IMPORTANCE



HYPERPARAMETER OPTIMIZATION

- GridSearch and Random Search methods are applied respectively.
- The best values are;

max_depth	None
max_features	auto
min_samples_split	2
n_estimators	500

RESULT

- Mean: 99
- Standard deviation: 30
- RMSE: 13.59