Sprint 1
Liquor Store Sales Prediction

Lucia Zou

Overview

Objective:

Using machine learning, the project aims to predict future sales, empowering store owners to maximize revenue and optimize operations.

Problem Statement:

Can we use machine learning models to predict future sales of liquor?

Why this dataset?

Solve the Problem

Step1: Get the Data (Kaggle)

Step2: Data Cleaning

(Fill in NA, duplicates, formatting)

Step3: EDA (Distributions, correlations)

Step4: Feature Engineering

Step5: Modeling

Potential Impact

1. Optimizing inventory and supply chains
2. Promote sales and revenue by having suitable marketing strategies
3. Satisfy more customers and enhance their overall experience

Introduction

The Dataset contains Iowa class E liquor holders data from Jan 2021 to Jan 2022. It has high potential with a lot of data to dig in. It contains 2805307 rows and 24 columns, 14 numeric and 9 categorical.

0	invoice_and_item_number	object
1	date	object
2	store_number	int64
3	store_name	object
4	address	object
5	city	object
6	zip_code	float64
7	store_location	object
8	county_number	float64
9	county	object
10	category	float64
11	category_name	object
12	vendor_number	float64
13	vendor_name	object
14	item_number	int64
15	item_description	object
16	pack	int64
17	bottle_volume_ml	int64
18	state_bottle_cost	float64
19	state_bottle_retail	float64
20	bottles_sold	int64
21	sale_dollars	float64
22	volume_sold_liters	float64
23	volume_sold_gallons	float64

dtypes: float64(9), int64(5), object(10)

Concerns

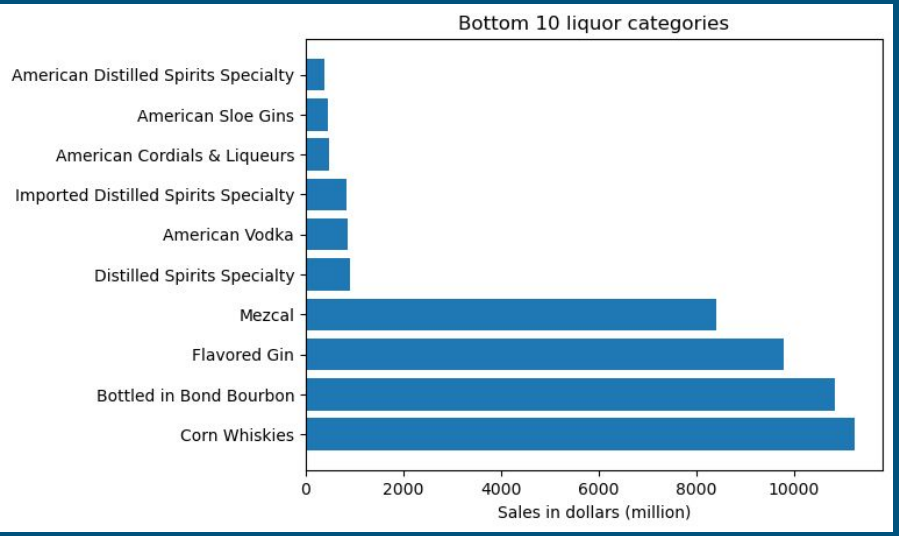
Limited to only Iowa state and only 13 months

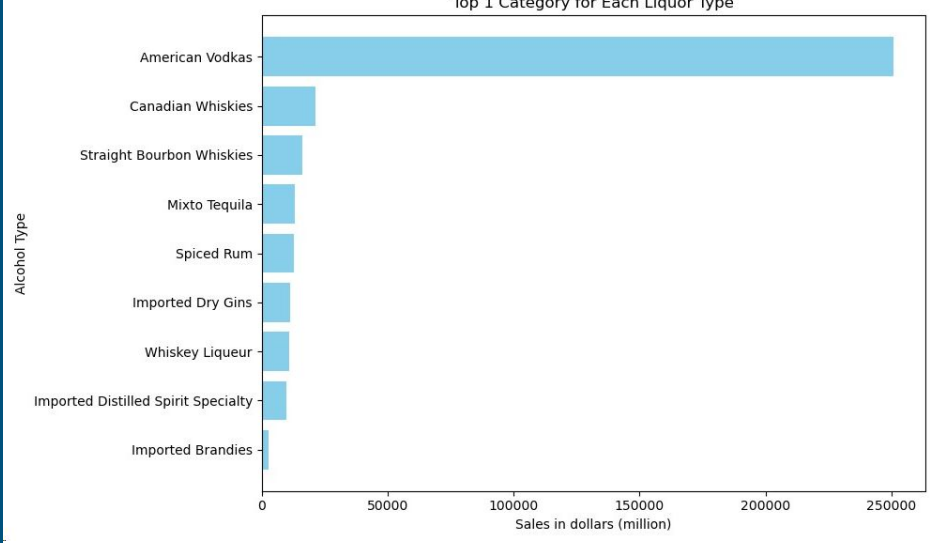
Findings

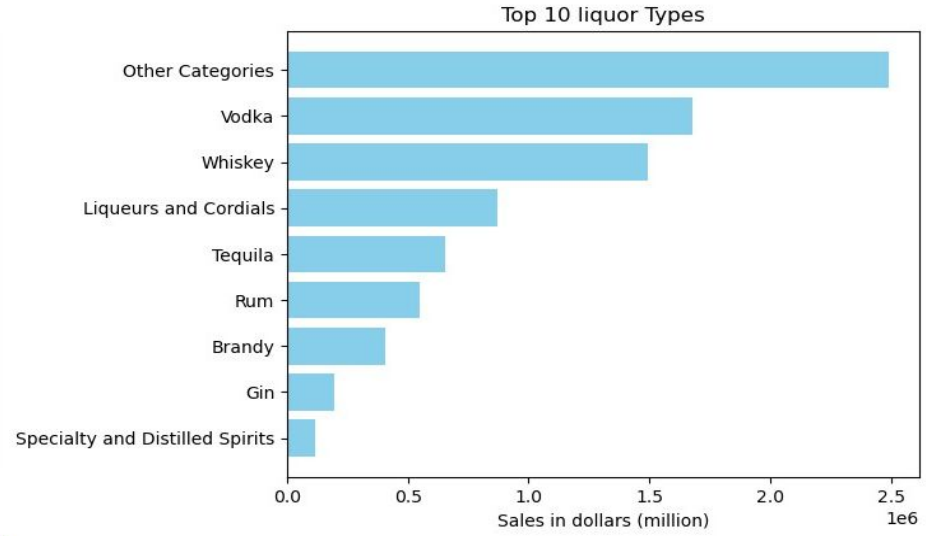
Average Bottles:12

Most popular type: Vodka

bottles_sold
2805307.000000
11.858369
35.668169
1.000000
3.000000
6.000000
12.000000
13200.000000



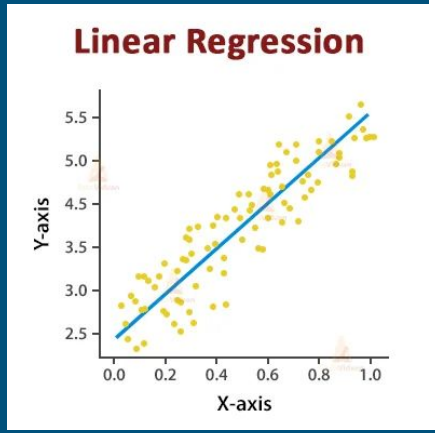


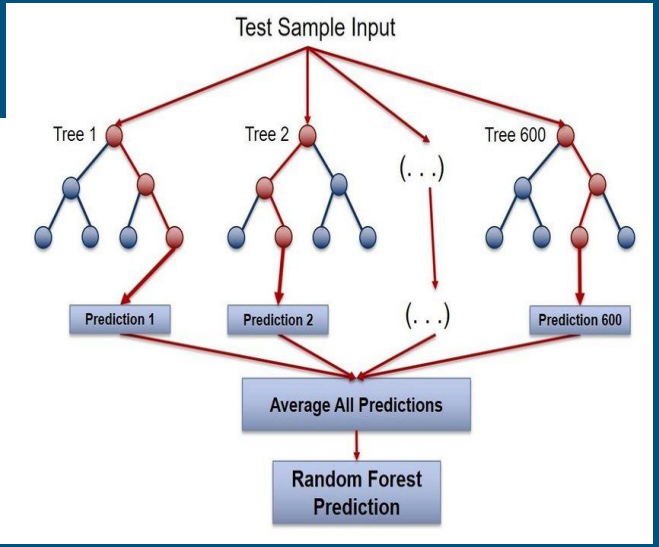


Furthermore...

Feature Engineering: change volume col into categorical

Basic Models: Regression





volume_sold_liters
0.1
10.5
10.5
1.5
0.5

Questions?

Thank you for listening!