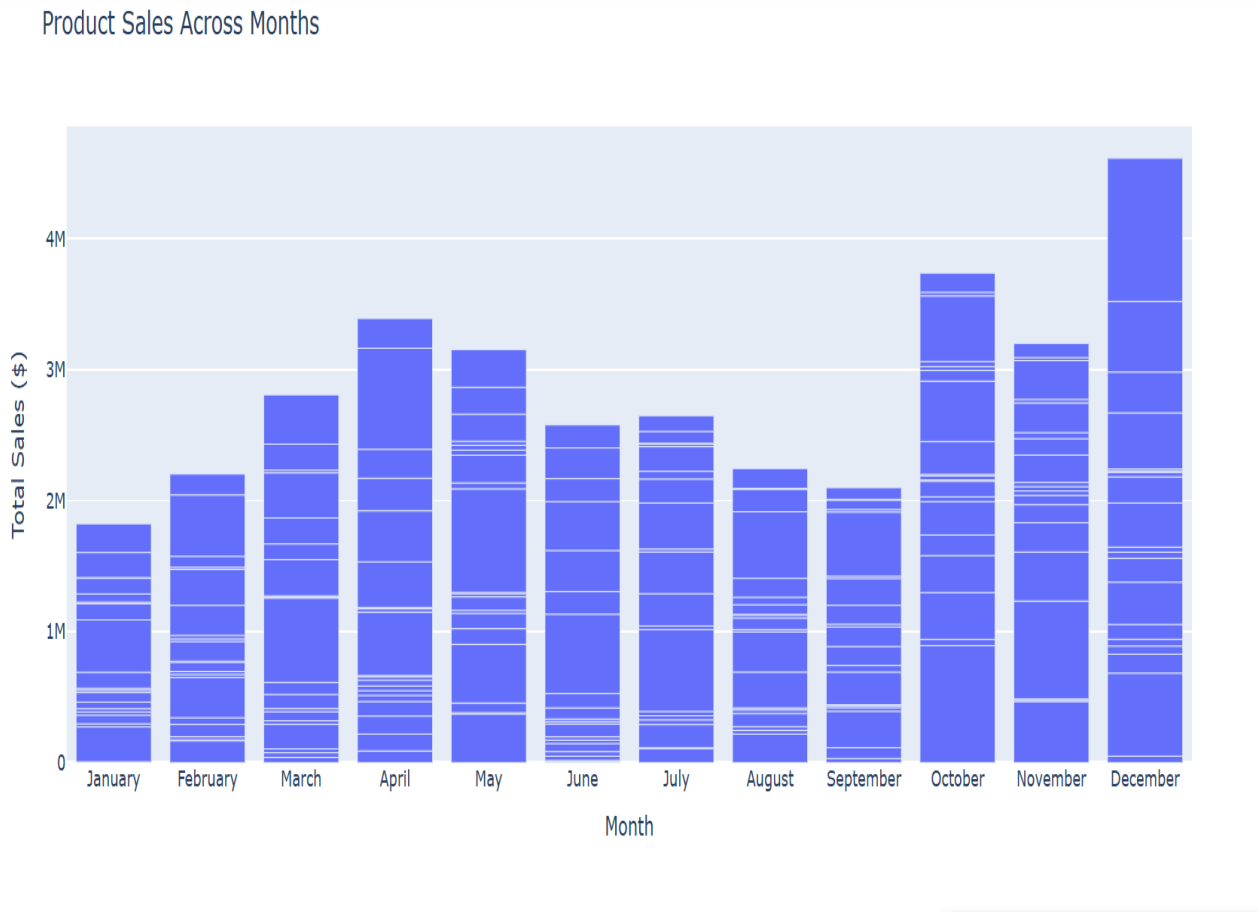
INSIGHTS, ANALYTICS, AND DEPLOYMENT ON ELECTRONICS SALES AND DIABETES DATA

Yash Shah

REVENUE INCREASE: PROMOTIONAL OFFERS SHOULD BE GIVEN IN Q1 AND Q3 TO INCREASE THE SALES



36.4 $ M worth of 19 different products sold across the year in 10 pin codes to 140 k unique customers

Q4 constitutes to 1/3rd of the sales for the year

Apple products contributes to 45% of sales ~15 M

CUSTOMER BASE INCREASE: PORTLAND CITY AMOUNTS TO ONLY 7% OF OVERALL CUSTOMERS WHICH FURTHER CAN BE TARGETED TO INCREASE CUSTOMER PENETRATION

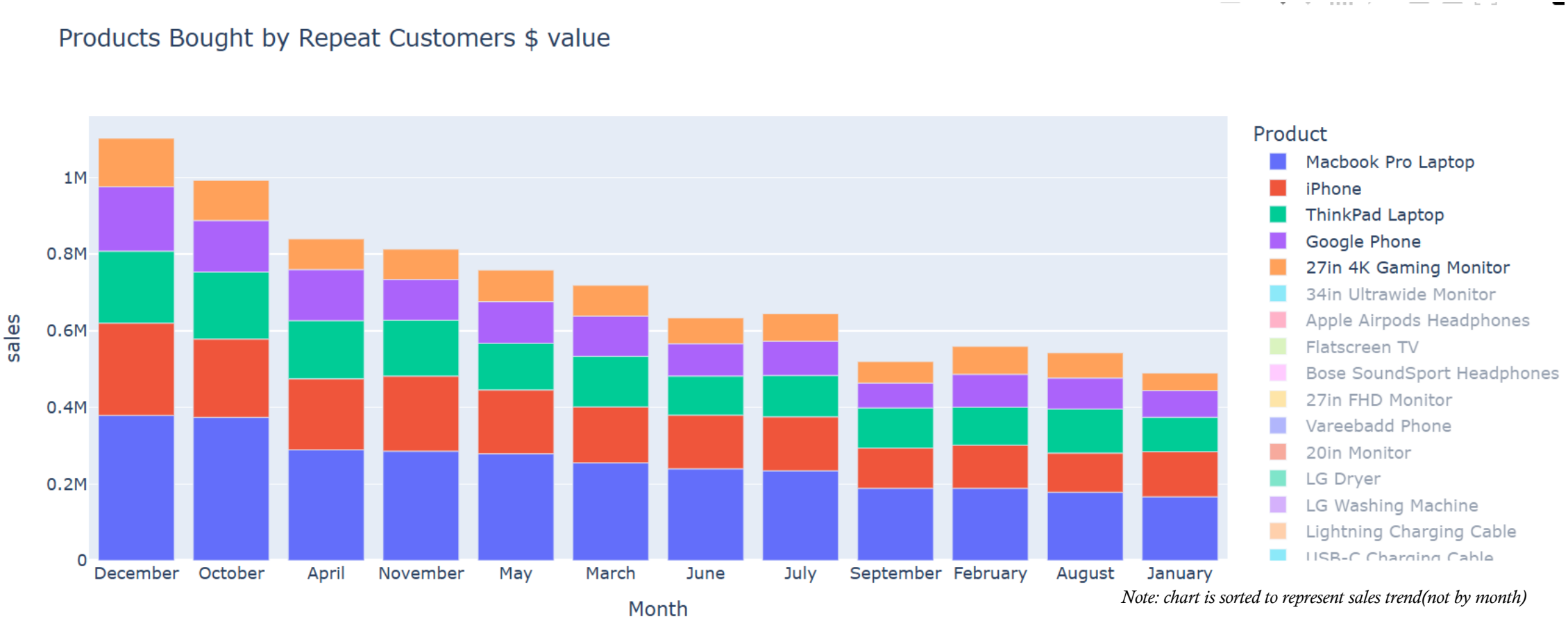
Postal Code	City	No of Customers	Percentage
WA 98101	Seattle	12212	8.67
CA 94016	San Francisco	28324	20.12
ME 04101	Portland	2301	1.63
OR 97035	Portland	8723	6.20
NY 10001	New York City	18807	13.36
CA 90001	Los Angeles	21450	15.24
TX 75001	Dallas	12321	8.75
MA 02215	Boston	15706	11.16
TX 73301	Austin	8609	6.11
GA 30301	Atlanta	12334	8.76

**note: unique customers are calculated based on unique addresses*

- Los angeles , New York and San Francisco combined brings 50 % of the customers
- Portland have customers in 2 postal code areas still amounts to only 7 % of customers bringing opportunity to Customer targeting specially in ME - 04101 postal code

CROSS SELLING/UP SELLING: BUNDLING I-PHONE WITH AIR-PODS, LIGHTNING CABLE TOGETHER HELPS IN GROWING REVENUE BY INCREASING AVERAGE ORDER VALUE INCREASE BY 20 %

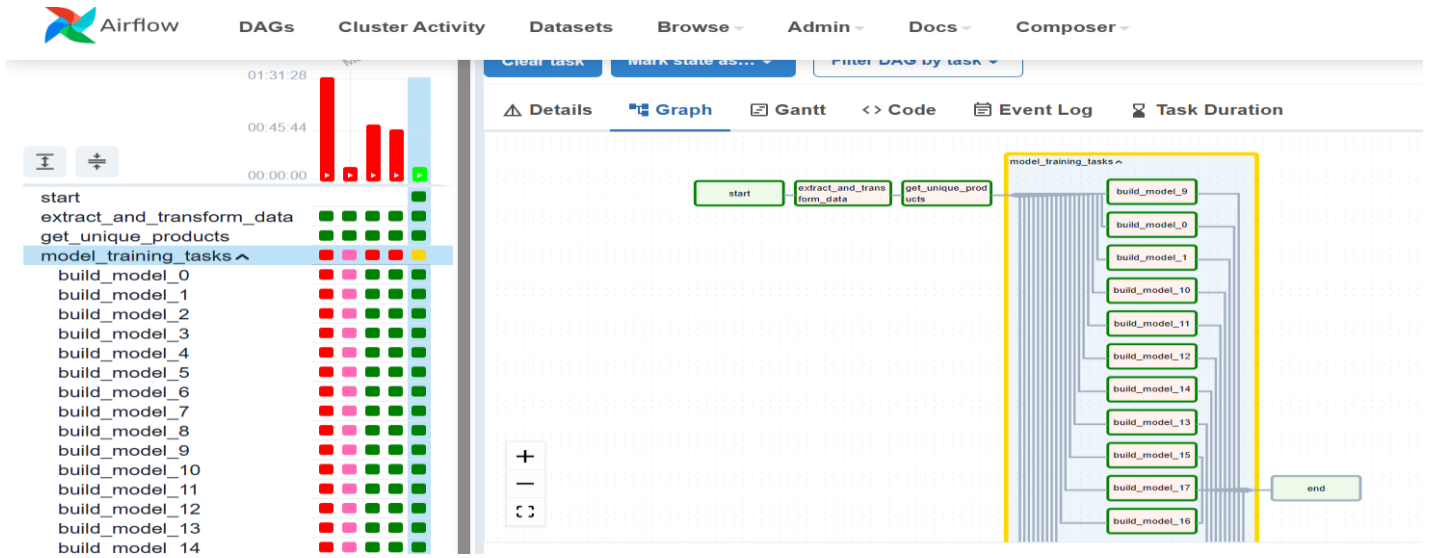
- 37 % customers repeat purchases every month
- Cross-Selling offers for existing customers - high ticket item like Macbook to in Q1 and Q4



Note: chart is sorted to represent sales trend(not by month)

STATE OF THE ART ML MODELS DEVELOPED AND DEPLOYED(AIRFLOW, GCP) TO PREDICT SALES FOR ALL PRODUCTS

- Scalable pipeline deployed for parallel training for all Products
- GCS enhouses the model artifects data

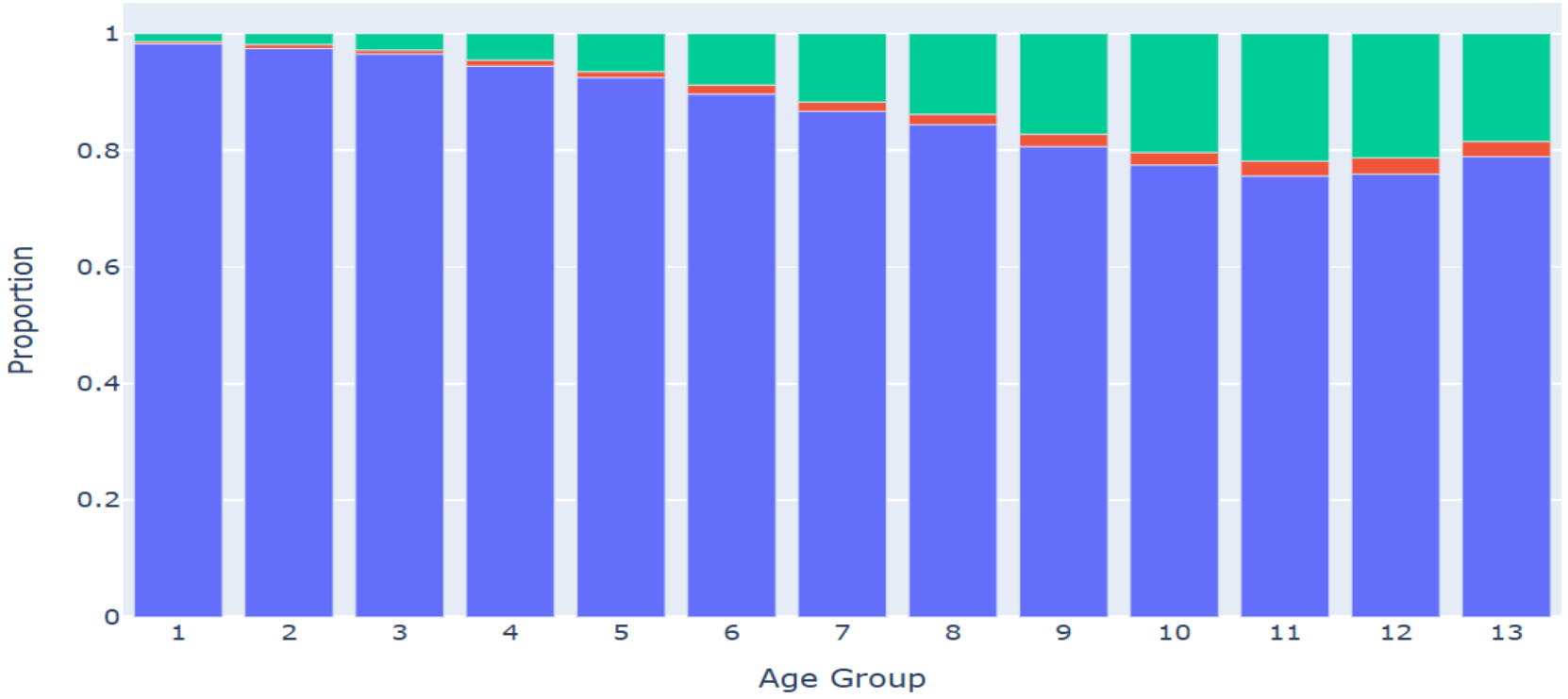


Next Steps: add github actions

Current Architecture

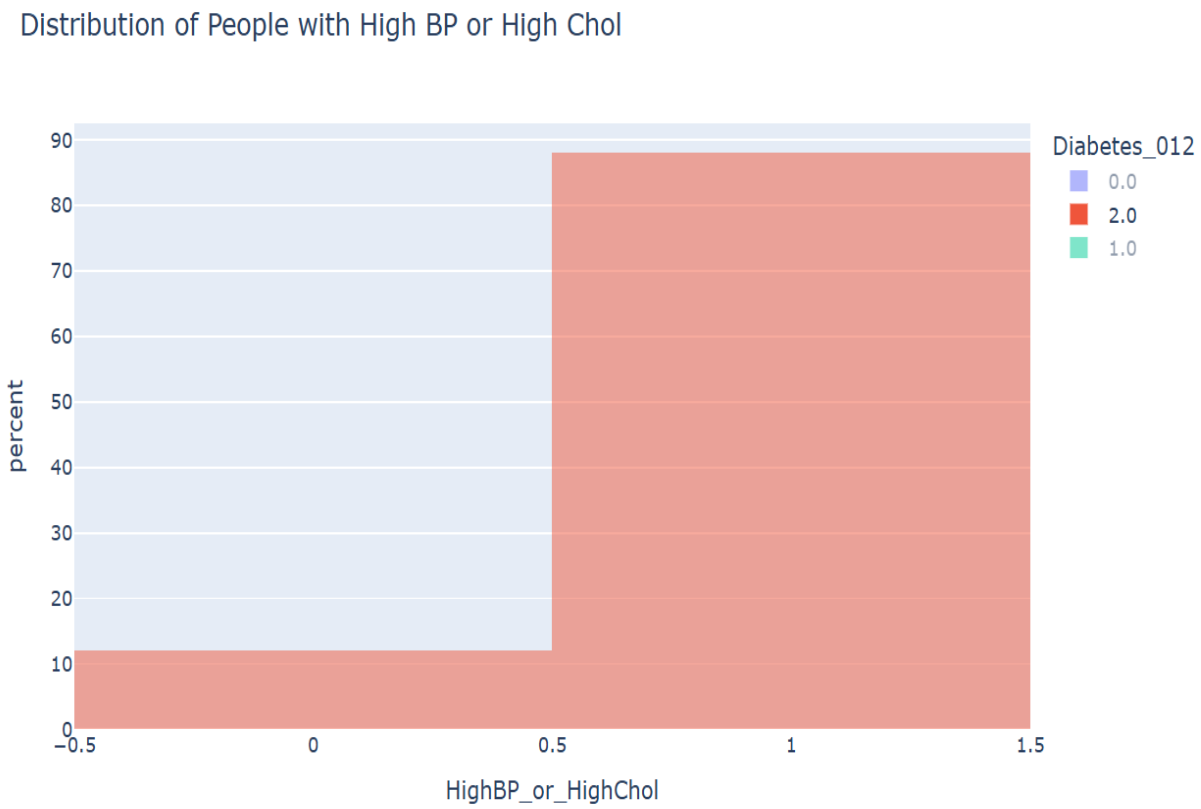
DIABETES ANALYTICS

90 % OF DIABETIC'S PATIENTS ARE AGE GROUP 7 OR HIGHER, WHERE AS AGE IS DIRECTLY PROPORTIONATE TO HAVING DIABETES



- Diabetes_Status
- = 0 → Normal
- = 1 → Pre-diabetic
- = 2 → Diabetics

MORE THAN 87% OF DIEBETIC PATIENTS HAVE EITHER HIGH CHOLESTEROL OR HIGH BLOOD PRESSURE



- 2 out of 3 diabetic patients have high cholesterol
- 75 % of them have high blood pressure

TOP 5 RISK FACTORS ASSOCIATED WITH DIABETES ARE FOUND BASED ON STATISTICAL AND MACHINE LEARNING METHODS

Top 5 Risk Factors

- BMI
- age
- Income
- Physical health
- General health

Methodology

- Stats Model - logit models and randomforest models are employed to find the top variables
- Based on p value and feature importance values

Higher BMI Linked to Diabetes:

in both studies independently and even After Controlling for Smoking and Physical Activity

- BMI is statistically significant in normal, pre-diabetic and diabetic people

```
# Bivariate analysis: ANOVA for continuous BMI
# here bmi have 3 groups so - this oneway stats check
f_stat, p_val = f_oneway(df_d[df_d['Diabetes_012'] == 0]['BMI'],
                         df_d[df_d['Diabetes_012'] == 1]['BMI'],
                         df_d[df_d['Diabetes_012'] == 2]['BMI'])
print(f"ANOVA F-statistic: {f_stat}, p-value: {p_val}")
```

```
ANOVA F-statistic: 6768.361066999288, p-value: 0.0
```

- BMI is statistically significant after controlling for Smoking status and physical activity

```
for i in logreg.coef_:
  print("BMI coefficients are" ,(i[0]))
  print("Smoker coefficients are" ,i[1])
  print("PhysActivity coefficients are" ,i[2])
```

```
BMI coefficients are -0.0465695953955307
Smoker coefficients are -0.18743490401056712
PhysActivity coefficients are 0.2948677621617713
BMI coefficients are 0.0164903142677311
Smoker coefficients are 0.04248754520343147
PhysActivity coefficients are -0.059486463884268495
BMI coefficients are 0.030079281127364993
Smoker coefficients are 0.1449473588070771
PhysActivity coefficients are -0.23538129827736456
```

APPENDIX

- More insights and charts can be found in notebooks