Boyota Customer Promotion plan

METIS BOOTCAMP

PLAN

Xinyuan Shu

2022/06/14

Boyota.CORP
METIS BOOTCAMP

Contents

CONTENTS

01.**Background introduction**

02. **Methodo -logy**

03. **Data Explan -ation & EDA**

04. **ML Models comparison**

05. **Deeper explore**

06.**Further Steps**

METIS®

Background Introduction

Business Problems Raised

- Boyota Corp. is a Japanese automobile company has plans to enter US markets with their existing products (P1, P2, P3, P4 and P5)
- Sales team has classified all customers into **A,B,C,D** class, they plan to use the same strategy on new markets and have identified 2627 new potential customers.
- We are required to help the manager to predict the right group of the new customers.

Methodology

Solution Path

Modeling Goal

Find Common/Different Characteristics of Customers:

- Gender
- Age
- Graduation
- Family Size
- Profession
- Marriage

……

ML Method

Find Suitable Machine Learning Methods

- KNN
- Random Forest
- Decision Tree
- Naïve Bayes

……

Business Goal

Correctly Assign Potential Customers to Their Segments

Achieving Successful Customer Management

↓

Maximize profits

METIS

Data Explanation & EDA

Features Explanation

Variable	Definition
ID	Unique ID
Gender	Gender of the customer
Ever_Married	Marital status of the customer
Age	Age of the customer
Graduated	Is the customer a graduate
Profession	Profession of the customer
Work_Experience	Work Experience in years
Spending_Score	Spending score of the customer
Family_Size	Number of family members for the customer (including the customer)
Var_1	Anonymised Category for the customer
Segmentation	(target) Customer Segment of the customer

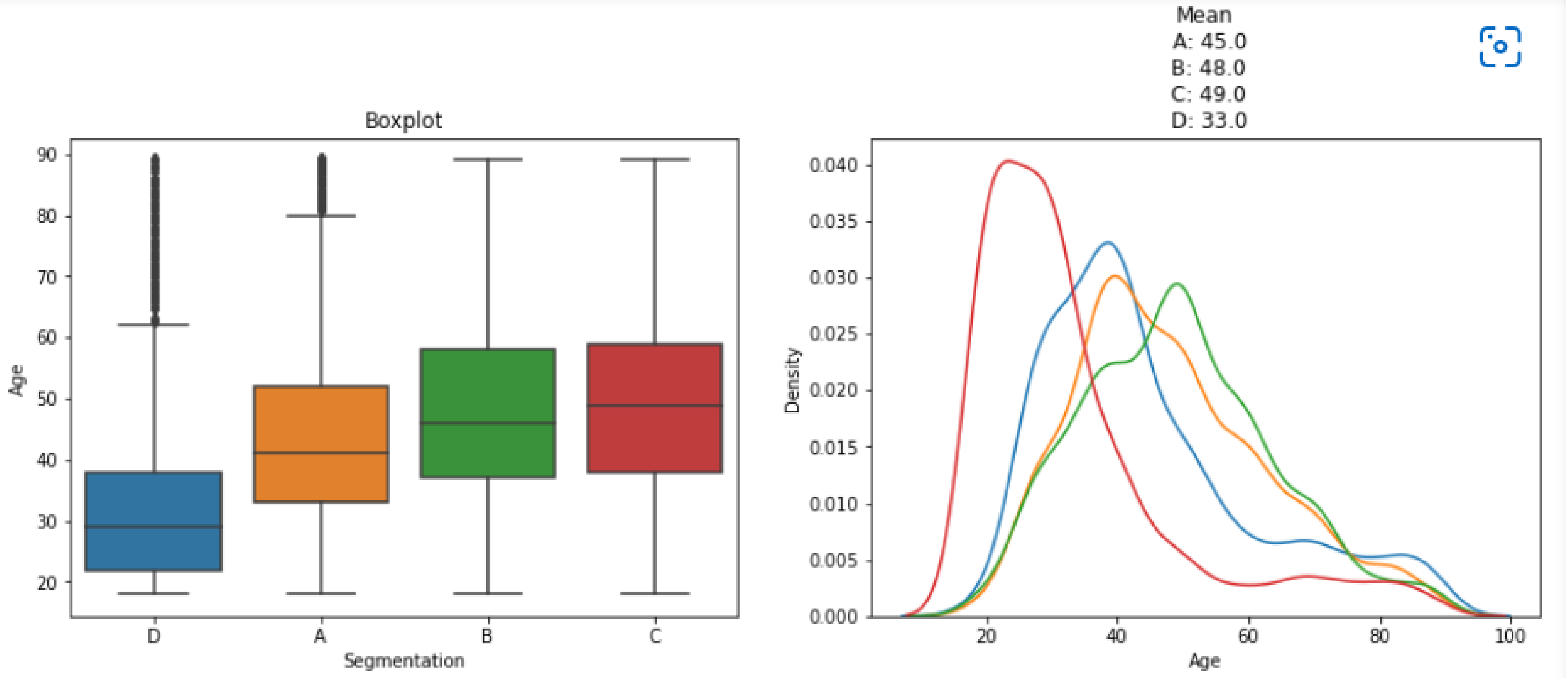
We will show some important Features:

- Work experience
- Spending Score
- Age
- Marriage
- Family Size

METIS

Exploratory Data Analysis-Important detection

Age

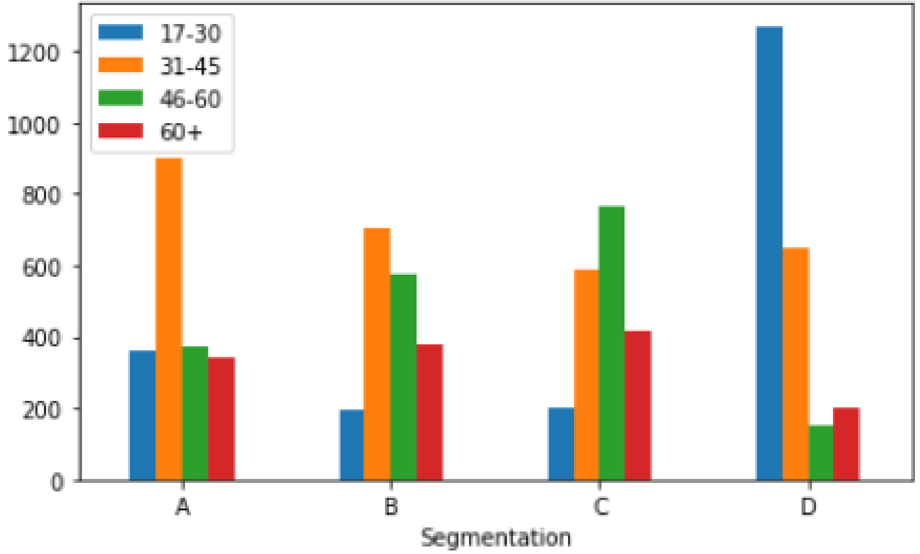


D: youngest C: oldest, D<A<B<C

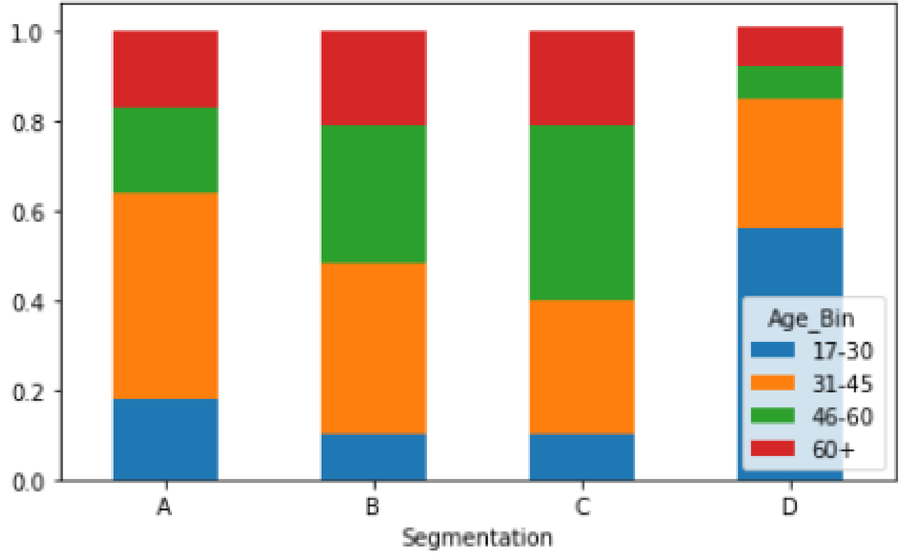
Exploratory Data Analysis-Important detection

Age-bin(17-30, 31-45,46-60,60+)

Segmentation	17-30	31-45	46-60	60+
A	360	898	373	341
B	195	707	575	381
C	201	588	767	414
D	1270	647	150	201



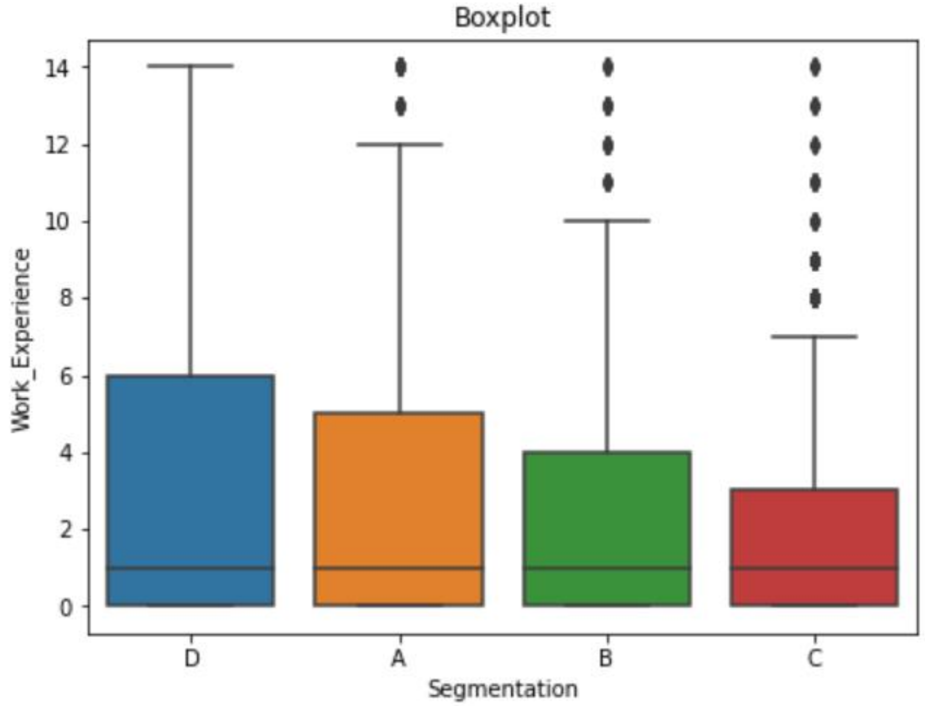
Age_Bin Segmentation	17-30	31-45	46-60	60+
A	0.18	0.46	0.19	0.17
B	0.10	0.38	0.31	0.21
C	0.10	0.30	0.39	0.21
D	0.56	0.29	0.07	0.09

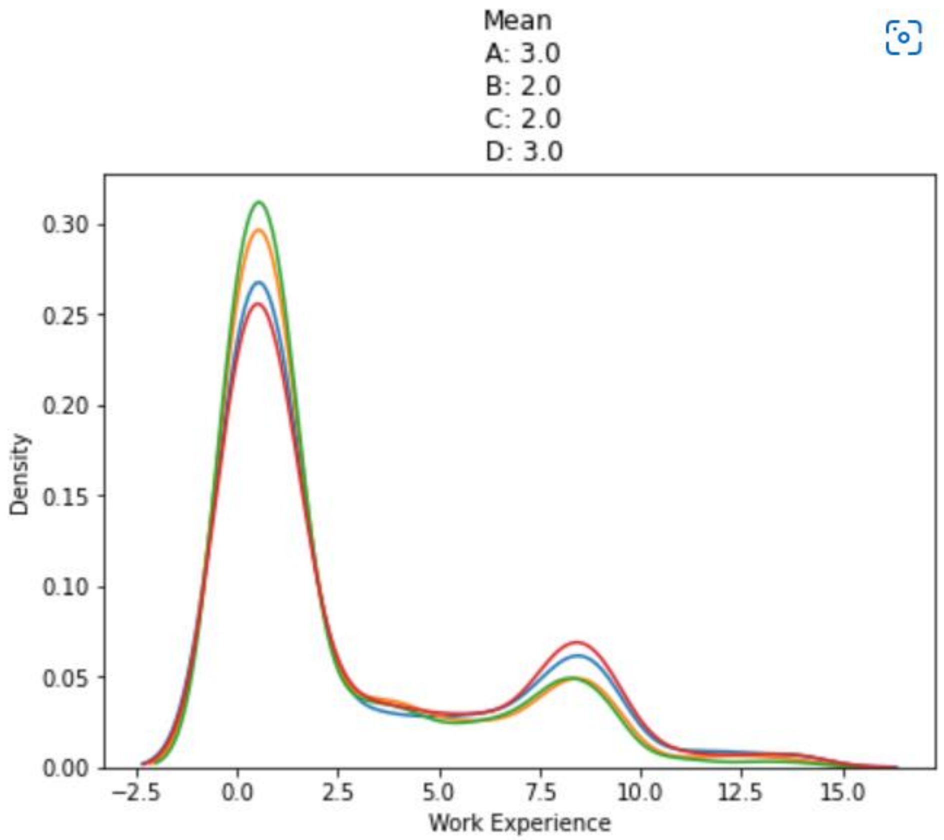


METIS®

Exploratory Data Analysis-Important detection

Work Experience



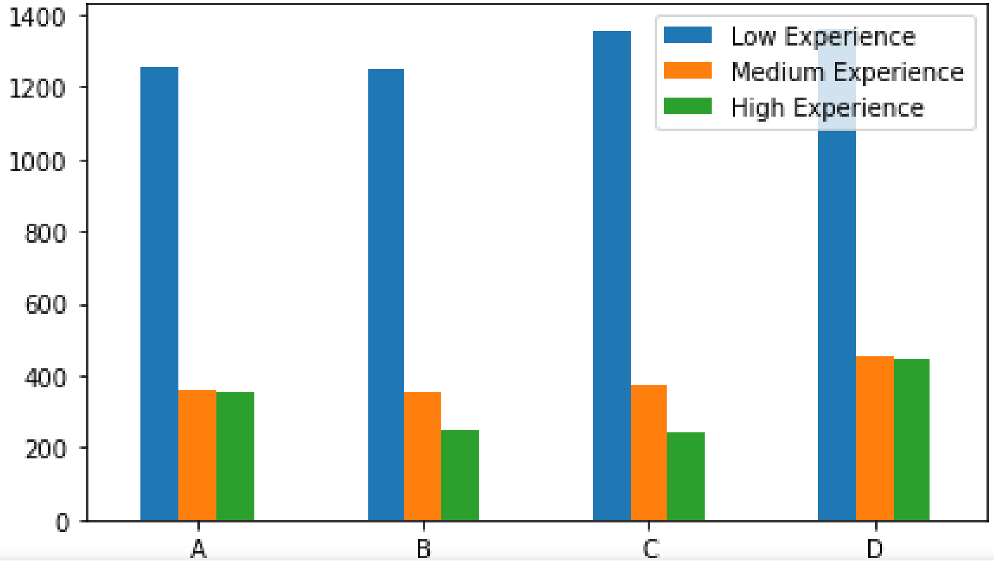


D: highest C: lowest, D>A>B>C

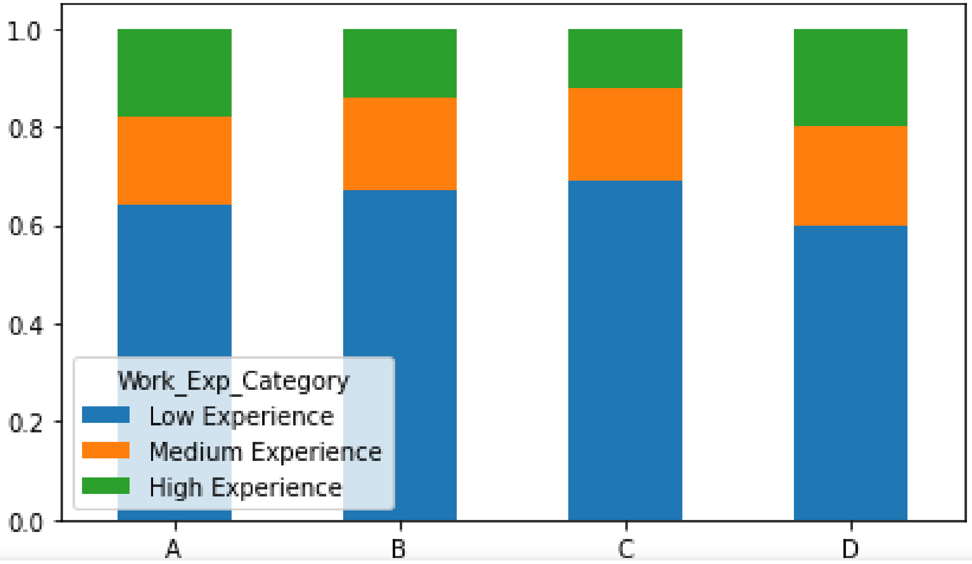
Exploratory Data Analysis-Important detection

Work Experience-bin(0-1,1-7,7-15)

Segmentation	Low Experience	Medium Experience	High Experience
A	1257	358	357
B	1252	354	252
C	1354	375	241
D	1363	456	449



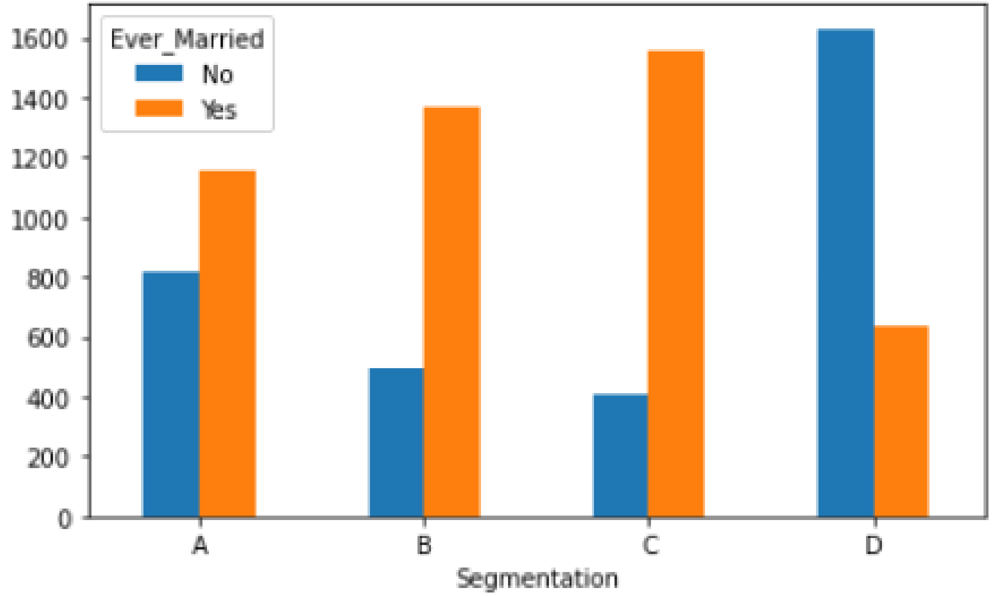
Work_Exp_Category Segmentation	Low Experience	Medium Experience	High Experience
A	0.64	0.18	0.18
B	0.67	0.19	0.14
C	0.69	0.19	0.12
D	0.60	0.20	0.20



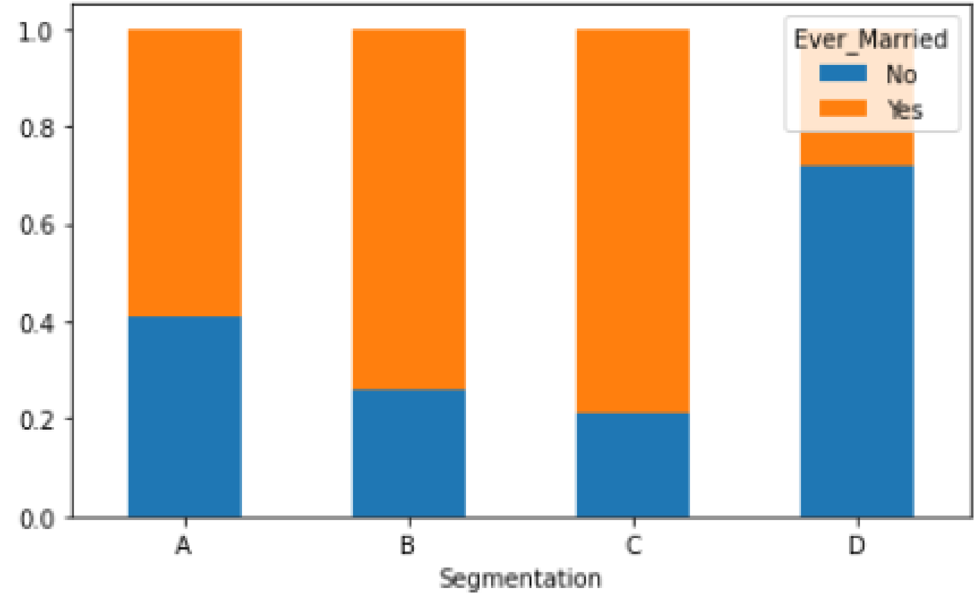
Exploratory Data Analysis-Important detection

Marriage

Ever_Married	No	Yes
Segmentation		
A	816	1156
B	491	1367
C	410	1560
D	1631	637



Ever_Married	No	Yes
Segmentation		
A	0.41	0.59
B	0.26	0.74
C	0.21	0.79
D	0.72	0.28

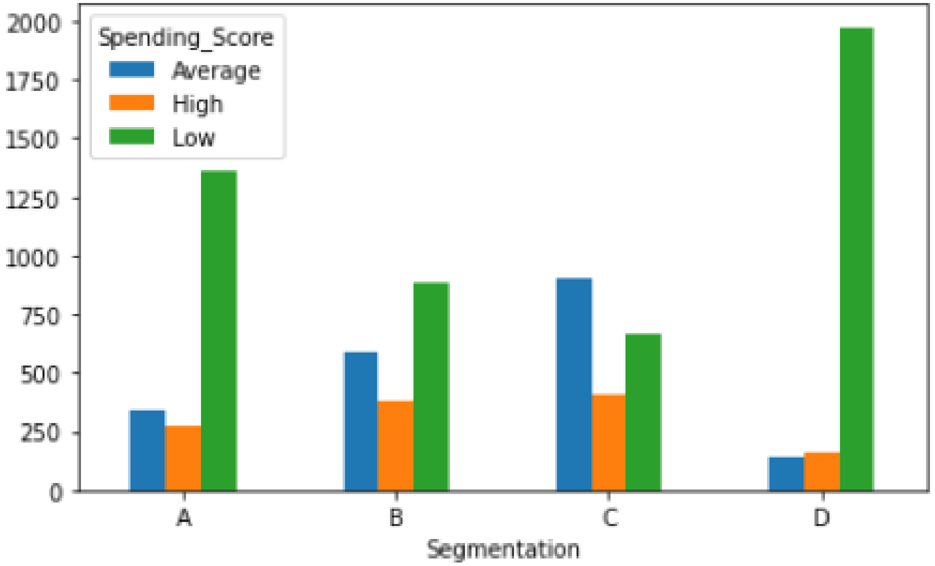


D: Most unmarried C: Most Married

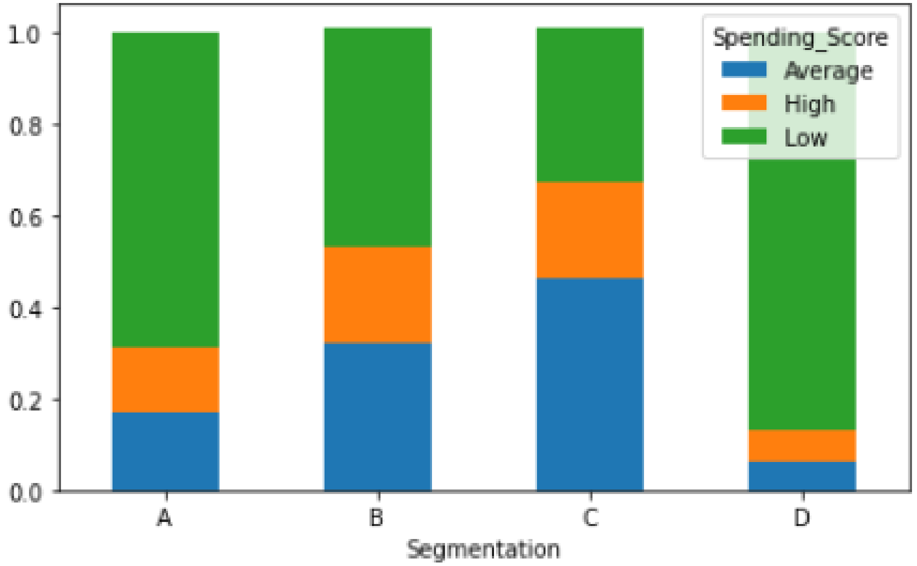
Exploratory Data Analysis-Important detection

Spending Score

Spending_Score Segmentation	Average	High	Low
A	343	271	1358
B	590	384	884
C	903	405	662
D	138	156	1974



Spending_Score Segmentation	Average	High	Low
A	0.17	0.14	0.69
B	0.32	0.21	0.48
C	0.46	0.21	0.34
D	0.06	0.07	0.87



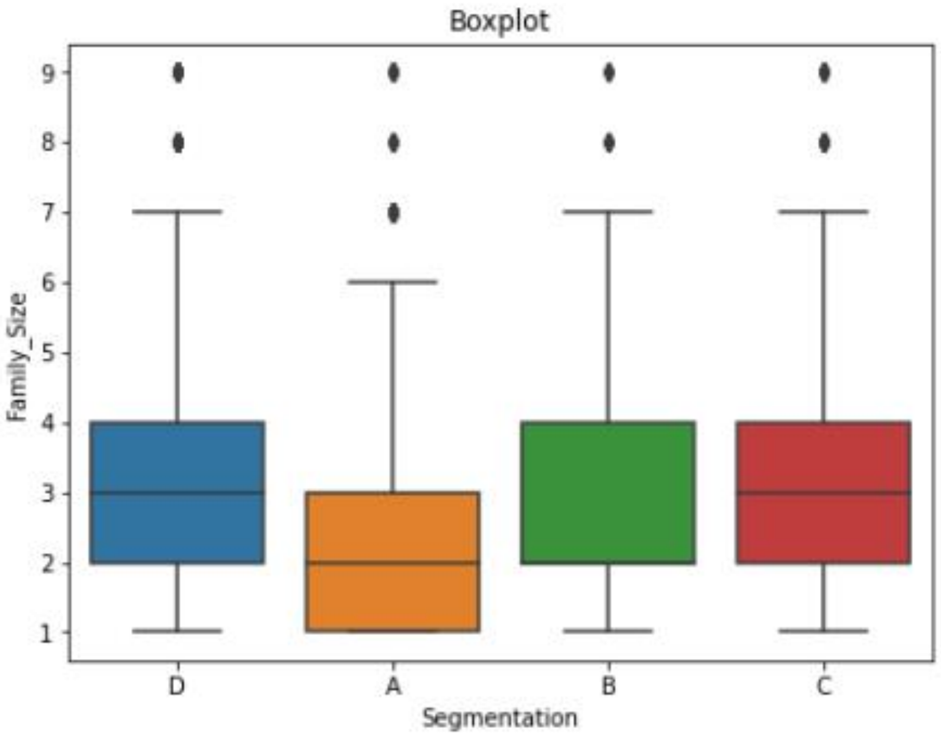
D: Most Low Spending Score C: Least Spending Score

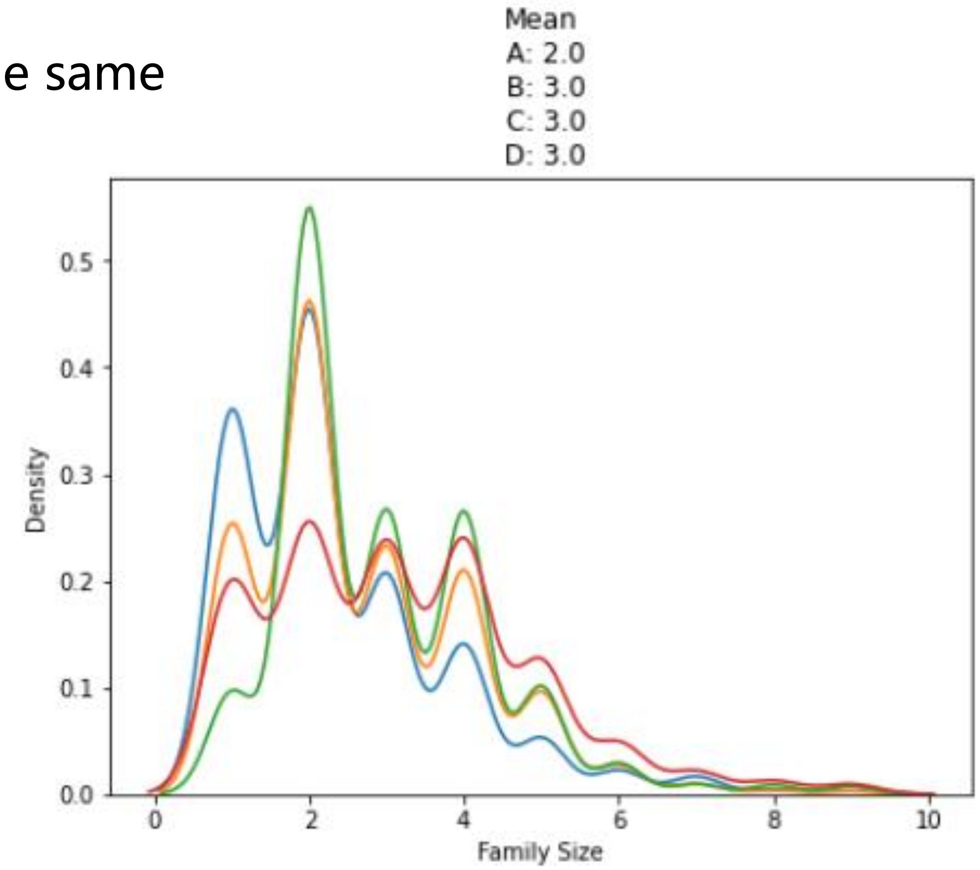
METIS

Exploratory Data Analysis-Important detection

Family Size

A: smallest Family size, Others about the same





Exploratory Data Analysis-Important detection

Conclusion

- Segment C (sketch: student/entry-level employees)
  - highest average age
  - married
  - lowest working experience
  - higher than other segments 'spending score
- Segment D (sketch: middle-aged wealthy customers)
  - lowest average age
  - unmarried
  - highest working experience
  - lower than other segments 'spending score
- Segment A has the smallest family size
- Hard to tell segments B from features

ML Models comparison

ML methods comparison

ML Methods	F1 Scores(Train)	F1 Scores(Test)
K-nearest neighbors(Normalized)	0.69	0.41
K-nearest neighbors(Dummy variables)	0.60	0.45
Random Forest(Normalized)	0.96	0.46
Random Forest(Dummy variables)	**0.72**	**0.47**
Naïve Bayes(Normalized)	0.48	0.47
Naïve Bayes(Dummy variables)	0.50	0.49

* I deleted dummy method of KNN and all Naïve Bayes method to simplify my Notebook

ML Models comparison

Final Choice of ML Methods

- Final Choice: Random Forest
  - Flexible, higher F1 scores
- Random Forest(Normalized)
  - Train data overfitting
- K-nearest neighbors
  -Both of F1 scores are lower than random forest.
  -In normalized dataset, KNN has a good performance may because that the original Dataset is built on K-means
- Naïve Bayes
  - Assumption not suitable

METIS®

Deeper explore

Segment B distinguish

	A	B	C	D
A	235	**148**	69	140
B	**155**	160	**153**	89
C	66	**145**	286	94
D	144	56	26	455

*Confusion Matrix of RF(Dummy method)

- 155 B are considered as A
- 148 A are considered as B
- 145 C are considered as B
- 153 B are considered as C
- B has the least obvious features as seen in EDA

Deeper explore

Segment B distinguish

	F1 Score(Train)	F1 Score(Test)
B	0.56	0.24
Not B	**0.91**	**0.85**

*Random Forest Methods For B/Not B Test

- Still hard to tell Segment B
- Easier to tell Segment which is not B

Conclusion

Conclusion& Inspiration

- From the EDA& Confusion Matrix:
  - Segment B is the hardest part to tell
  - Segment C & D have clearer customer sketches

- Reason for choosing Random Forest:
  - Flexibility
  - Ability to deal with noise
  - Ability to deal with more features
  - Nice F1 Score (and precision score)

- New models to tell not segment B groups

- Accuracy is low no matter what we choose

METIS

Further Steps

Further Steps to Impel Project

ID Features Explore

ID seems have some information to tell but we directly drop it

Web App

Build a web app by flask for better visualization and communication

Boyota CORP.

Thank You

Boyota Customer Promotion plan

METIS BOOTCAMP

Xinyuan Shu

2022/6/14