

PREDICTING CUSTOMER PRODUCT

Capstone Project Document



NGOC QUYNH MAI NGUYEN (MARY)

Data Scientist

29/05/2021

Contents

[Business background 2](#_Toc73187188)

[Business Opportunity 4](#_Toc73187189)

[Stakeholders 4](#_Toc73187190)

[Business question 4](#_Toc73187191)

[Data question 4](#_Toc73187192)

[Data 4](#_Toc73187193)

[Data science process 5](#_Toc73187194)

[Data analysis 5](#_Toc73187195)

[Modelling 5](#_Toc73187196)

[Outcomes 10](#_Toc73187197)

[Data answer 10](#_Toc73187198)

[Business answer 10](#_Toc73187199)

[References 11](#_Toc73187200)

# Business background

Hopnhat International Express – HNC was founded in 2001, with main services including courier and transportation. After 20 years of operation, HNC has demonstrated the good quality of services with competitive prices and confirmed leading brand in courier (express) and transportation in Vietnam.

Website: <https://hopnhat.com/>

Services:

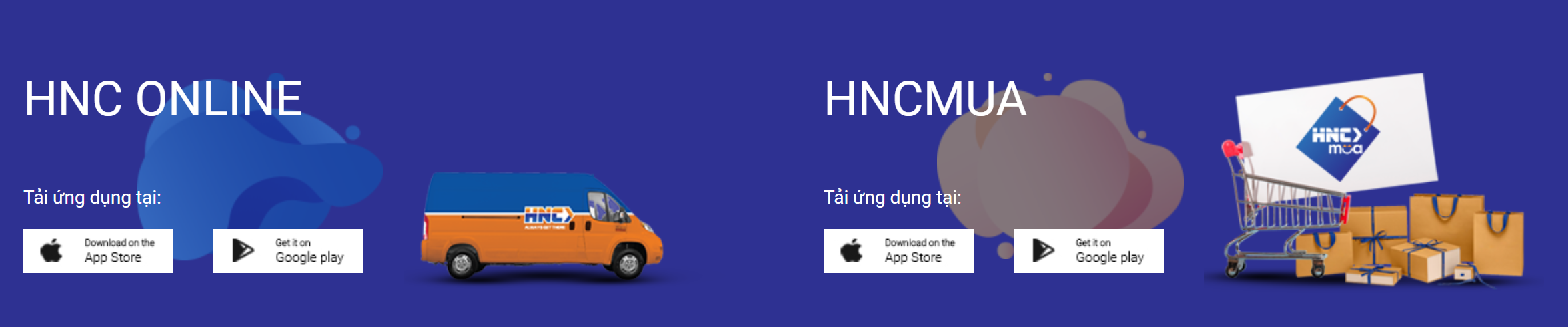
* Air freight
* Heavy cargo transportation
* Nationwide export consolidation services
* Cargo import services
* Ocean Freight
* International newspapers publishing service
* E-commerce
* Custom clearance service

Countries: more than 113 countries

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| No | Country | No | Country | No | Country |
| 1 | AUSTRALIA | 39 | ISRAEL | 77 | PORTUGAL |
| 2 | AUSTRIA | 40 | ITALY | 78 | QATAR |
| 3 | AZERBAIJAN | 41 | IVORY COAST | 79 | REPUBLIC OF SERBIA |
| 4 | BAHRAIN | 42 | JAPAN | 80 | ROMANIA |
| 5 | BANGLADESH | 43 | JORDAN | 81 | RUSSIA |
| 6 | BELGIUM | 44 | KAZAKHSTAN | 82 | RWANDA |
| 7 | BENIN | 45 | KENYA | 83 | SAN MARINO |
| 8 | BOTSWANA | 46 | KOREA SOUTH | 84 | SAUDI ARABIA |
| 9 | BULGARIA | 47 | KUWAIT | 85 | SIERRA LEONE |
| 10 | BURKINA FASO | 48 | LATVIA | 86 | SINGAPORE |
| 11 | BURUNDI | 49 | LEBANON | 87 | SLOVAKIA |
| 12 | CANADA | 50 | LESOTHO | 88 | SOMALIA |
| 13 | CHINA | 51 | LIBYA | 89 | SOMALILAND |
| 14 | CONGO | 52 | LIECHTENSTEIN | 90 | SOUTH AFRICA |
| 15 | CROATIA | 53 | LITHUANIA | 91 | SPAIN |
| 16 | CYPRUS | 54 | LUXEMBOURG | 92 | SRI LANKA |
| 17 | CZECH REPUBLIC | 55 | MACAU | 93 | SUDAN |
| 18 | DEM REP OF THE CONGO | 56 | MALAWI | 94 | SWAZILAND |
| 19 | DENMARK | 57 | MALAYSIA | 95 | SWEDEN |
| 20 | DJIBOUTI | 58 | MALI | 96 | SWITZERLAND |
| 21 | EGYPT | 59 | MALTA | 97 | SYRIA |
| 22 | ENGLAND | 60 | MAURITIUS | 98 | TAIWAN |
| 23 | ESTONIA | 61 | MOLDOVA | 99 | TANZANIA |
| 24 | ETHIOPIA | 62 | MONACO | 100 | THAILAND |
| 25 | FINLAND | 63 | MOROCCO | 101 | TOGO |
| 26 | FRANCE | 64 | MOZAMBIQUE | 102 | TUNISIA |
| 27 | GABON | 65 | MYANMAR | 103 | TURKEY |
| 28 | GEORGIA | 66 | NAMIBIA | 104 | UGANDA |
| 29 | GERMANY | 67 | NEPAL | 105 | UKRAINE |
| 30 | GHANA | 68 | NETHERLANDS | 106 | UNITED ARAB EMIRATES |
| 31 | GREECE | 69 | NEW ZEALAND | 107 | UNITED ARAB EMIRATES/ AUH |
| 32 | HONG KONG | 70 | NIGERIA | 108 | UNITED ARAB EMIRATES/ DXB |
| 33 | HUNGARY | 71 | NORWAY | 109 | UNITED STATES |
| 34 | INDIA | 72 | OMAN | 110 | UZBEKISTAN |
| 35 | INDONESIA | 73 | PAKISTAN | 111 | YEMEN |
| 36 | IRAN | 74 | PALESTINE | 112 | ZAMBIA |
| 37 | IRAQ | 75 | PHILIPPINES | 113 | ZIMBABWE |
| 38 | IRELAND | 76 | POLAND |  |  |

# Business Opportunity

* Developing e-commerce by selling product from overseas
* Developing air freight service by selling product in store for customer
* Website: <https://hncmua.com/>



# Stakeholders

* Stakeholders: CFO and Manager
* Stakeholders’ expectations: increase revenue by 40% per year

# Business question

* Main business question: What products customer want to buy?
* The business value of answering this question: Choosing the right products can increase revenue by 40% per year
* The required accuracy: above 95% - choosing the wrong products will lead to a bad debt ratio because of over stock

# Data question

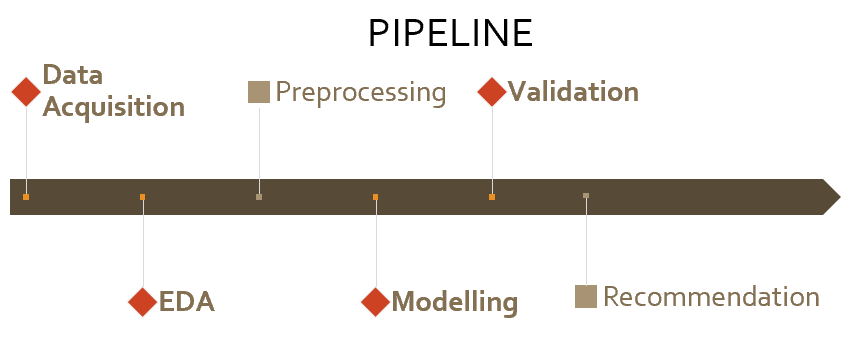
* The data question: Based on customer’s purchase and shipping history, how accurately can we predict the products that will be in the next order?
* The data required to answer the question: customer’s purchase and shipping history

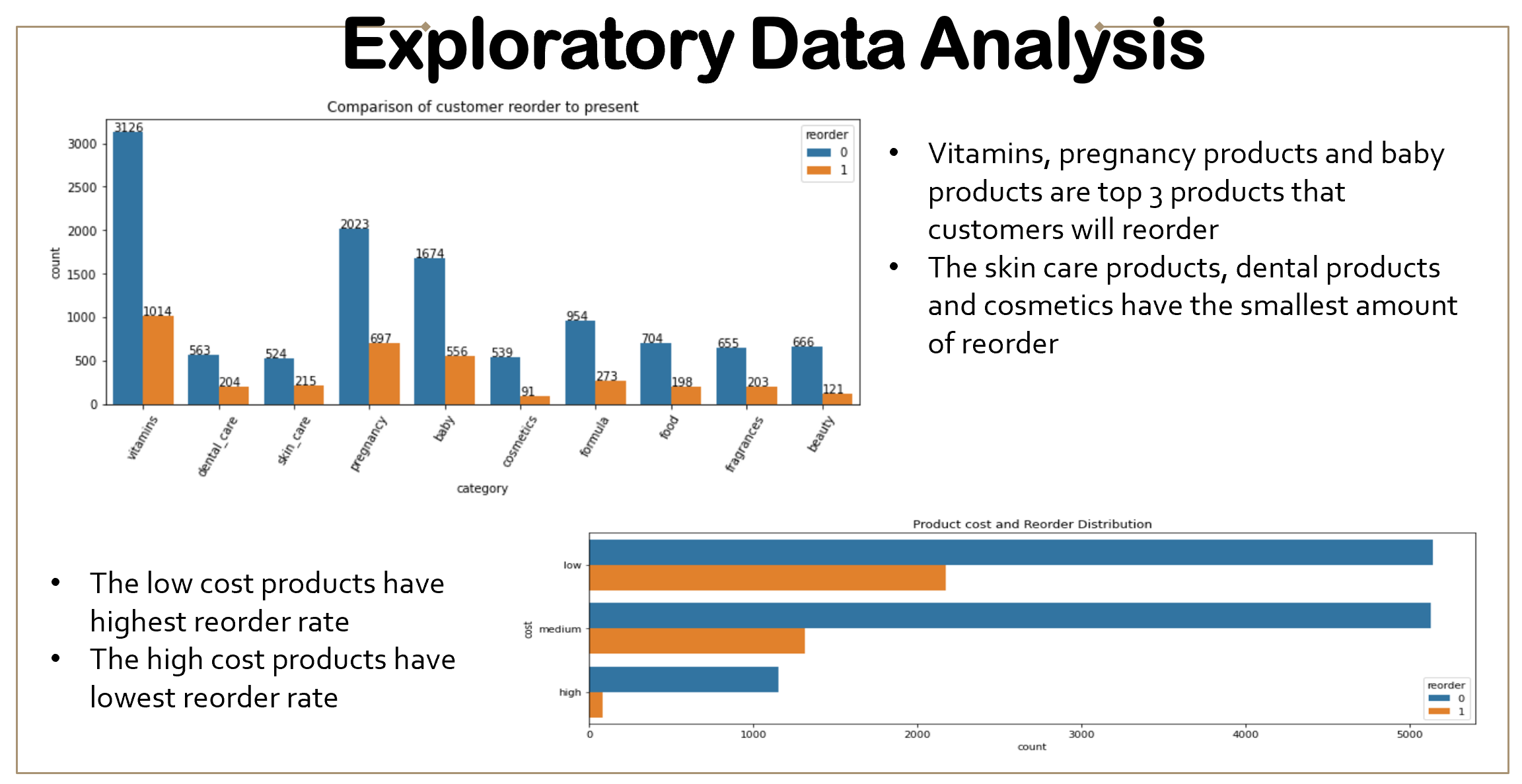
# Data

* The dataset was collected internally when customers shipping the products to other countries or when customers buying the products
* The dataset contains 15,000 observations
* The dataset was collected based on the real life customer orders
* The dataset does not have any missing values as all the features are the required fill to process order.

# Data science process

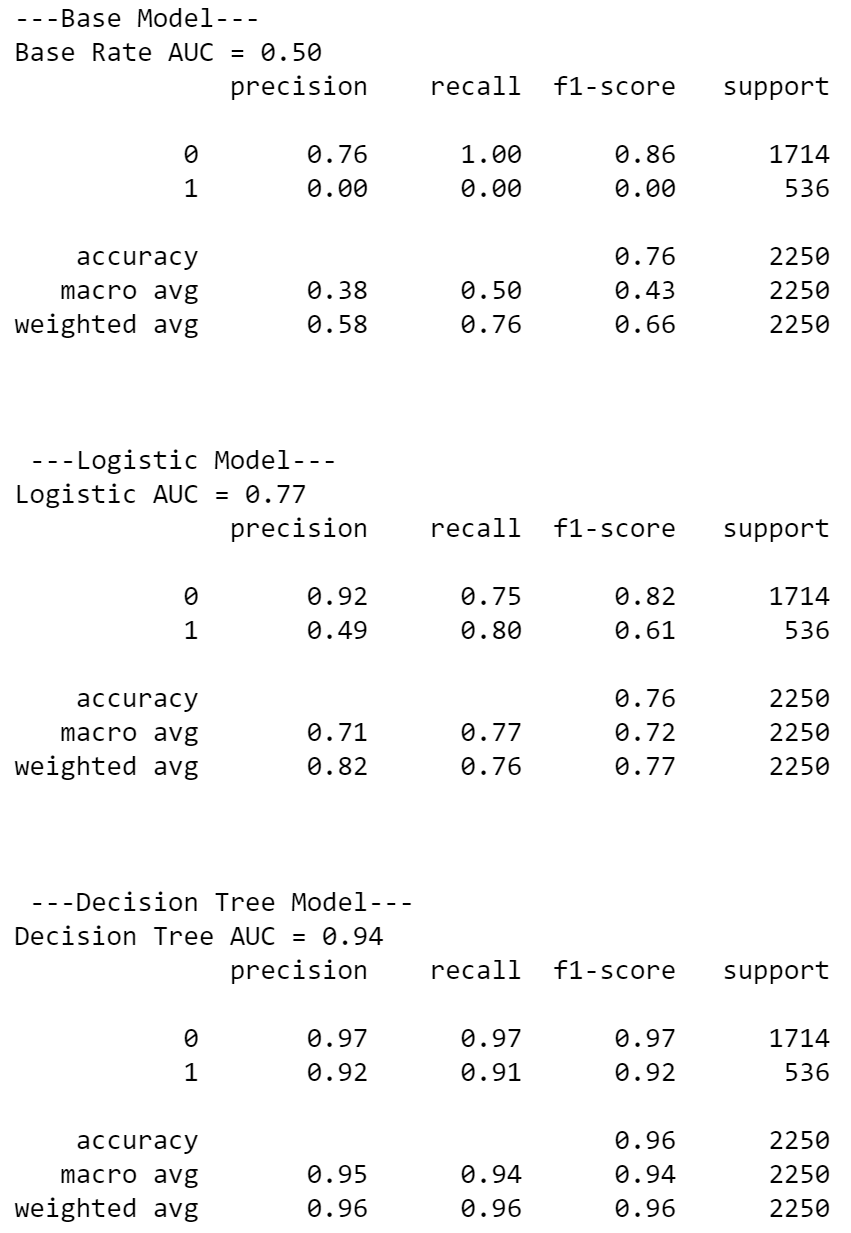
## Data analysis

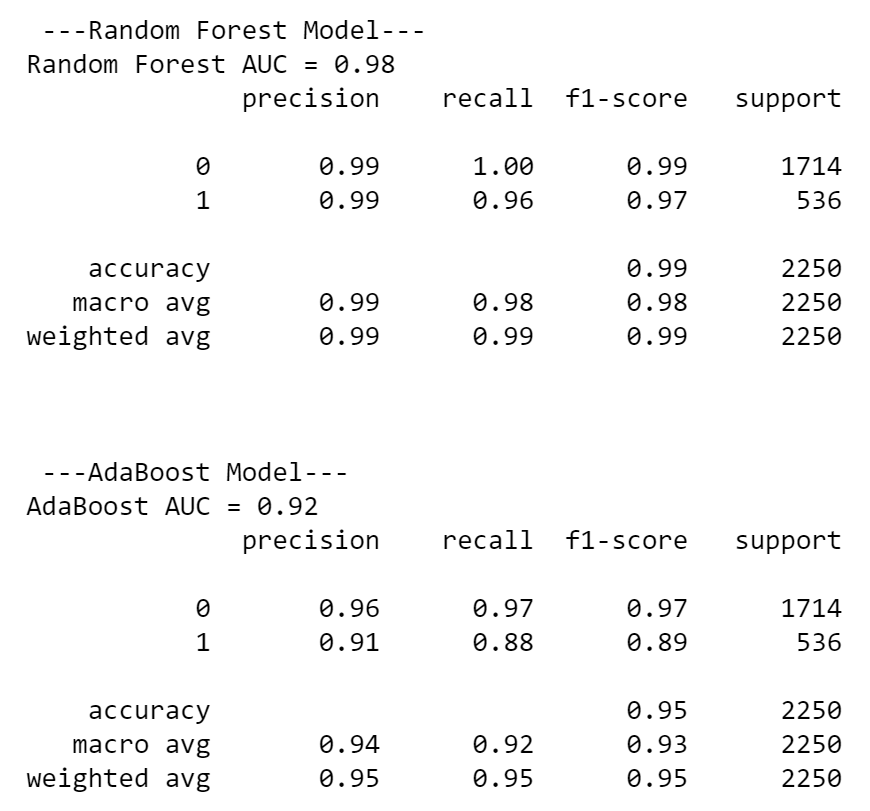


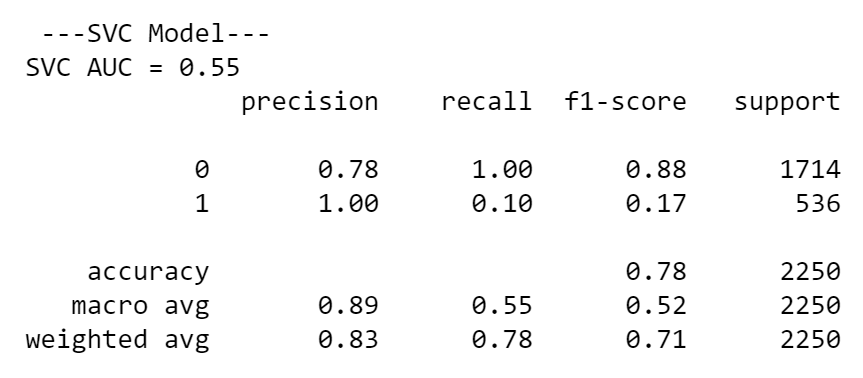


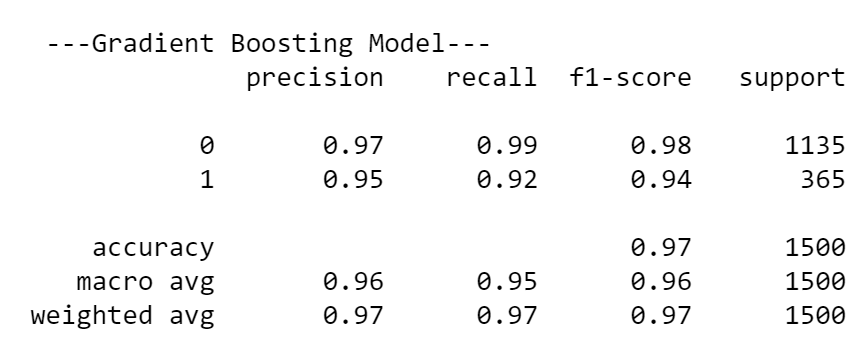
## Modelling

* The main features used: product category, reorder rate, customer satisfaction, discount rate, product cost
* Base Rate - Logistic Regression - Decision Tree - Random Forest - AdaBoost Model – SCV - Gradient Boosting

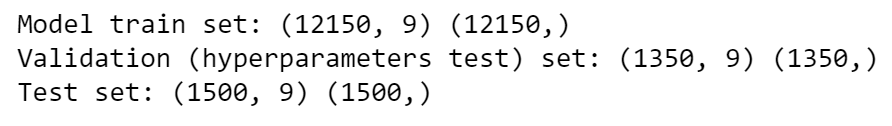


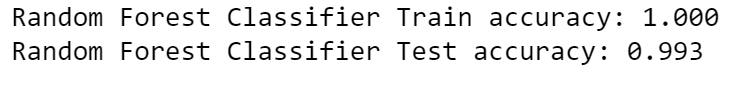


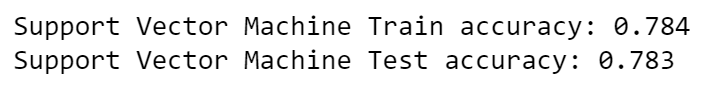


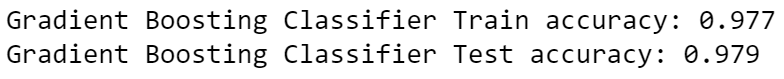


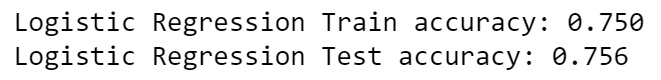
* Cross Validation

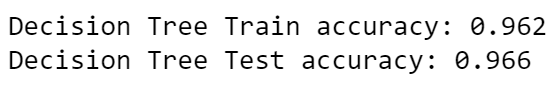


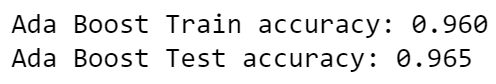


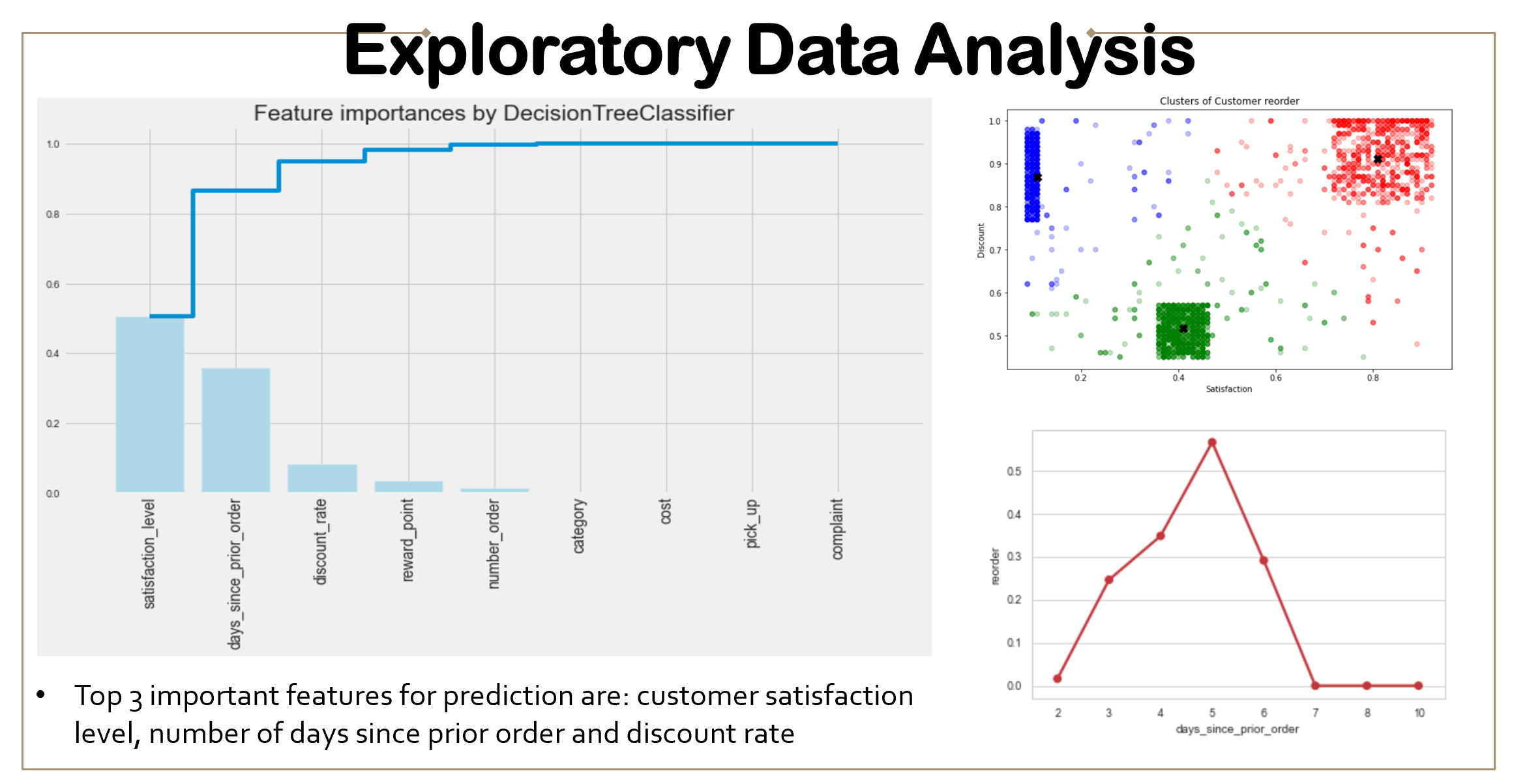


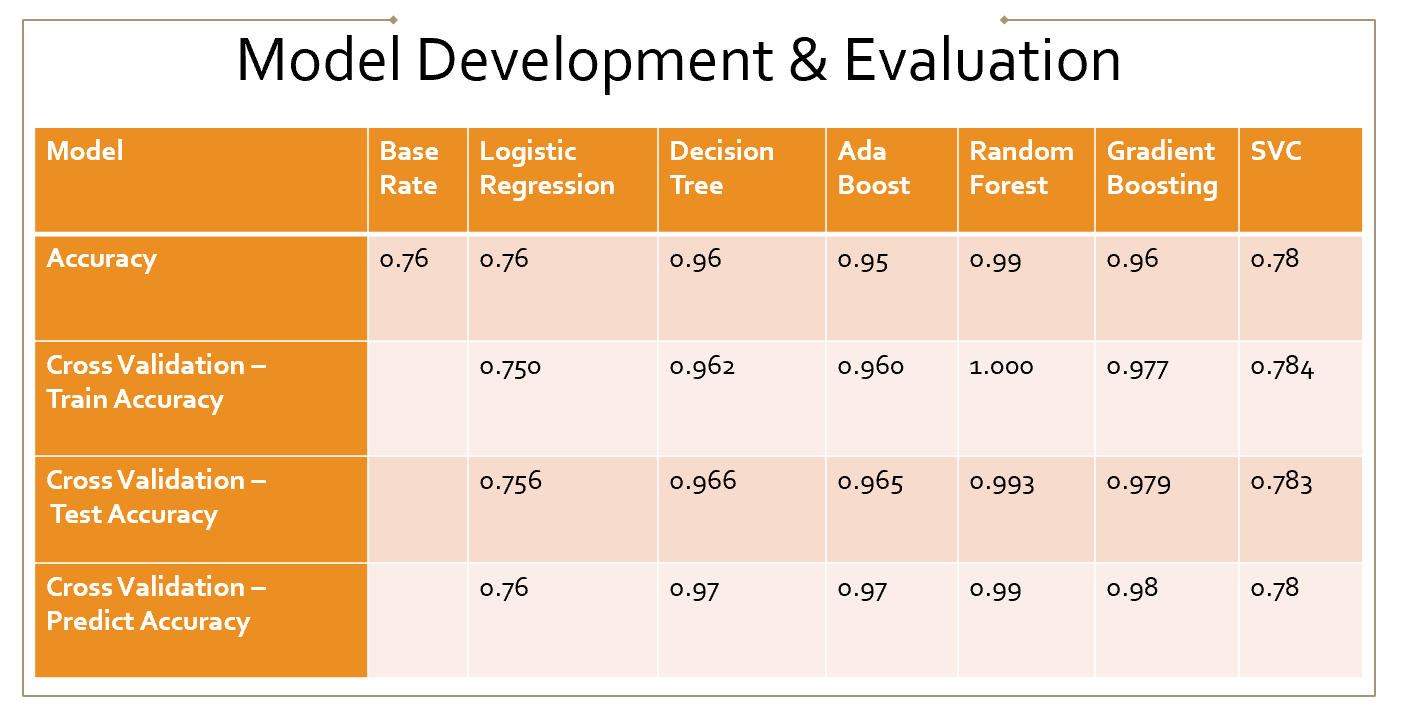


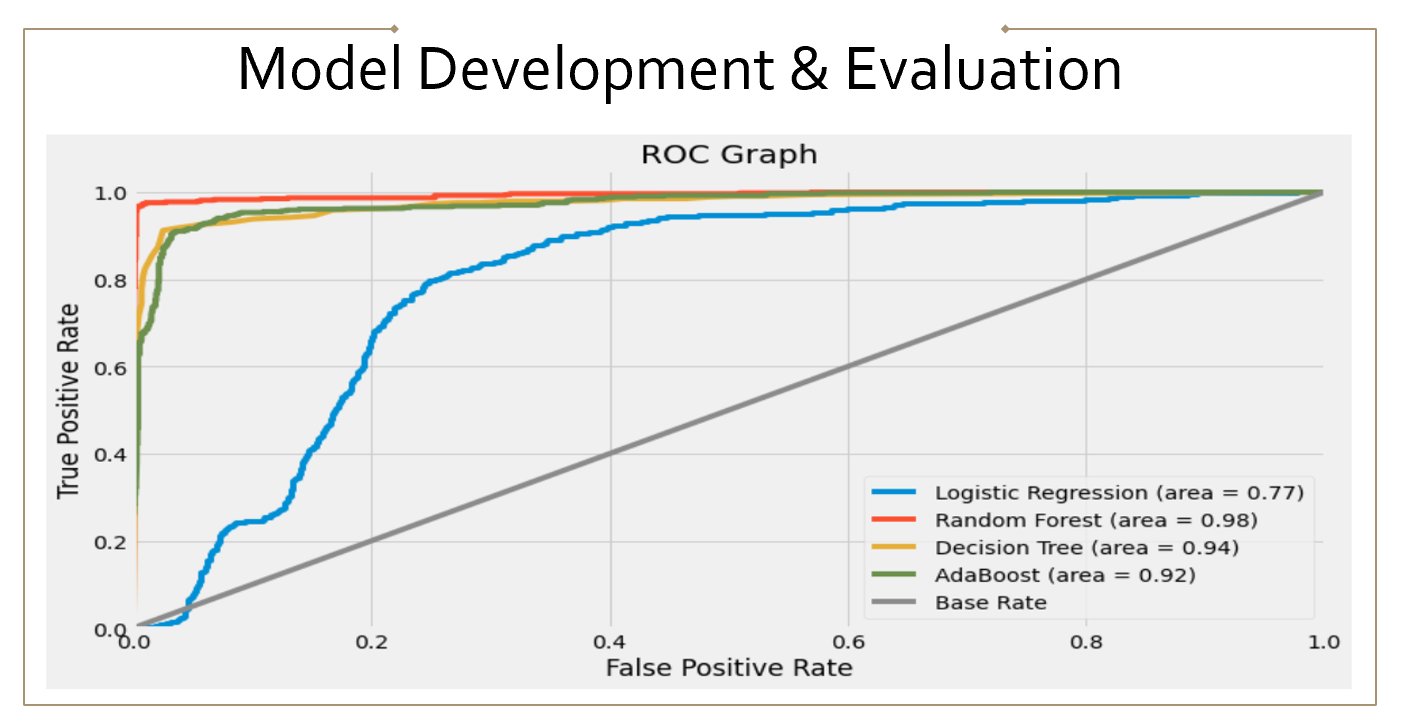












## Outcomes

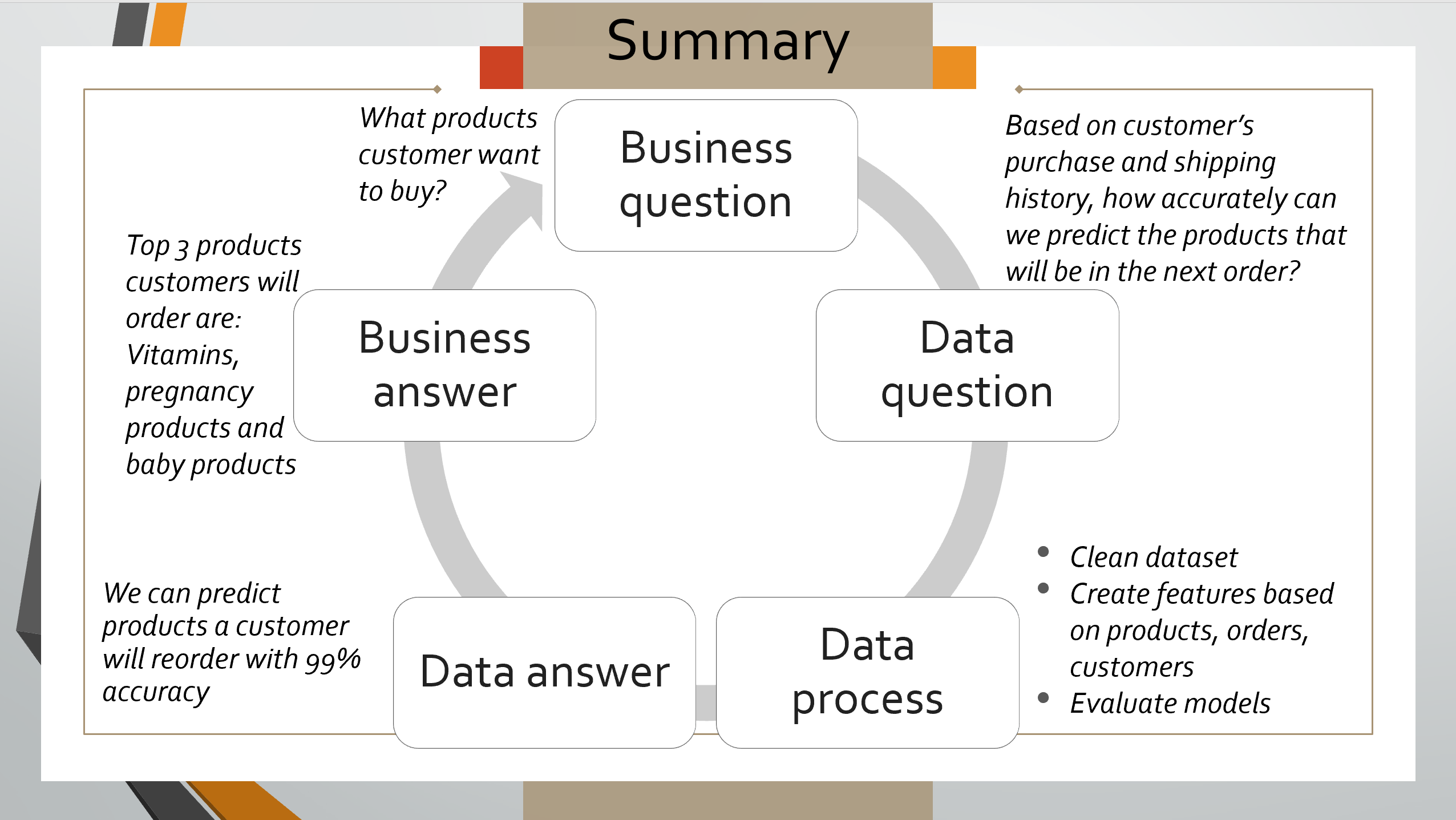
* Model selection: Random Forest Model as it has the highest accuracy rate – 99%

# Data answer

* We can predict products a customer will reorder with 99% accuracy

# Business answer

* Top 3 products customers will order are: Vitamins, pregnancy products and baby products
* The confidence level in the business answer: 95-99% accuracy



# References

* Dataset: collected internally from the company
* Resources:

<https://hopnhat.com/>

<https://hncmua.com/>

<https://github.com/MaryNguyen94/Institute-of-Data/tree/IOD>