Utilize web behavior metrics to enhance user purchase rate

BOS YUJIA FENG

2025/10/28

# Business Context & Problem Statement

- **ISSUE:**
  - The current **purchase rate** of the product is **15.48%,** but there is **still room to improve conversion efficiency**.
  - Personalized recommendation should be improve

- **GOAL:**
  - Use **PageValue and ExitRate** to identified High-value customers(high PageValue, low ExitRate).
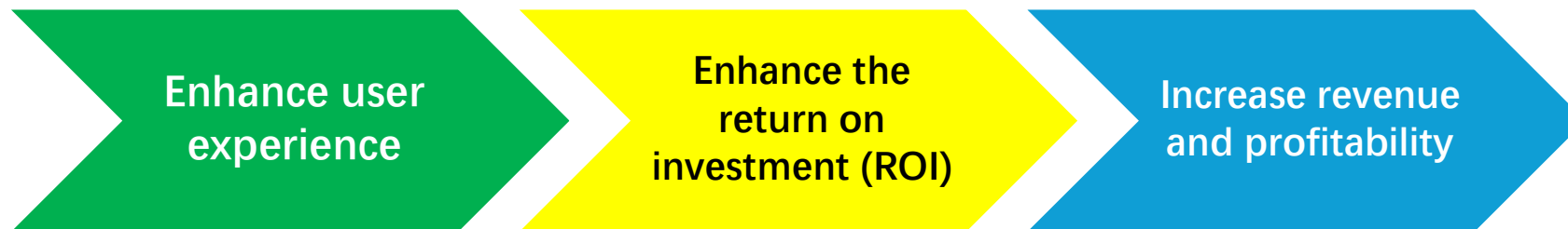 - Make more targeted decisions for **different customer types**

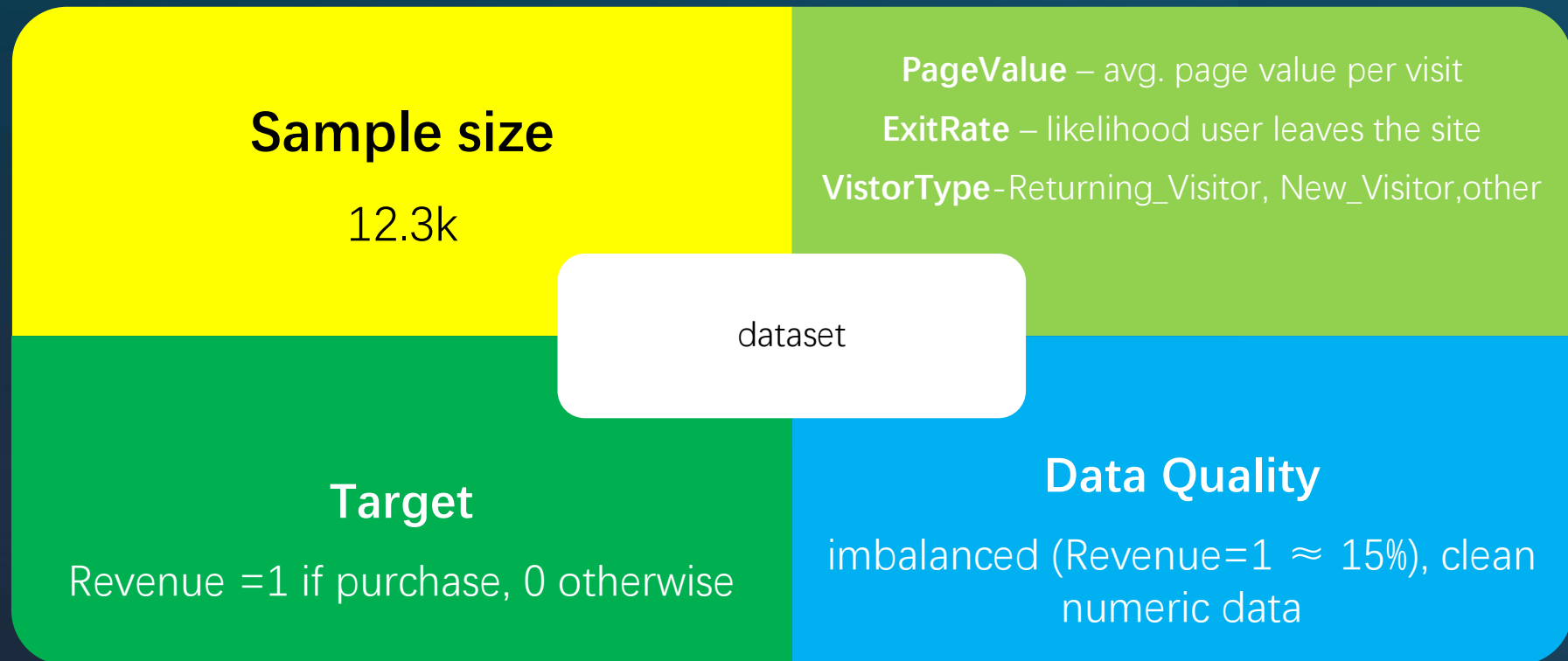**Revunue**

10423

1909

0   1
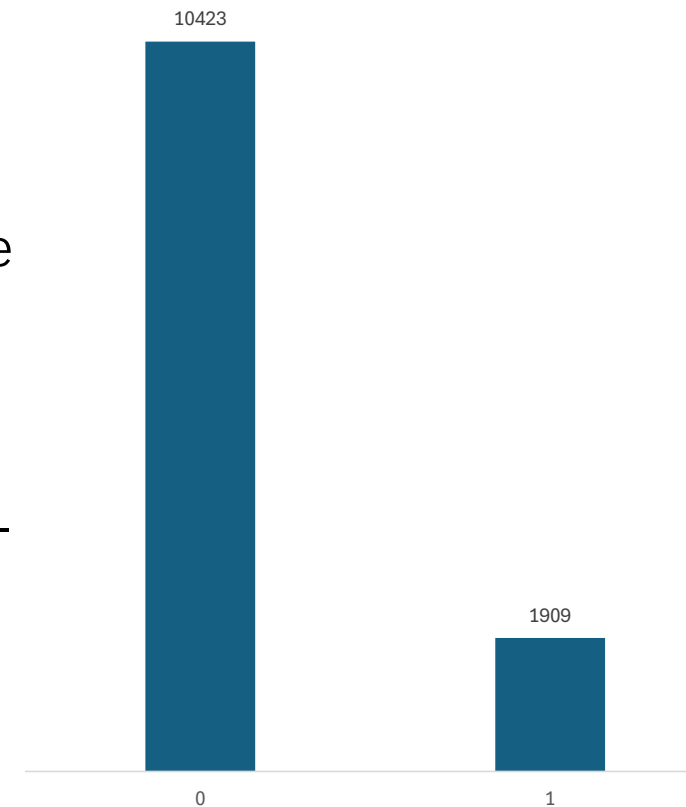
Why Webpage Behavior Metrics Matter for Online Sales

- **Online shopping and e-commerce are an important trend**
- **Precise advertisements and product pages** are indispensable parts of product sales
- **EXAMPLE:**The e-commerce company of Century Hearing Aids increased the conversion rate by 220%, the revenue by 300%, and the profit by 3000%.

Enhance user experience → Enhance the return on investment (ROI) → Increase revenue and profitability

Dataset Overview

Sample size

12.3k

PageValue – avg. page value per visit

ExitRate – likelihood user leaves the site

VistorType - Returning_Visitor, New_Visitor, other

dataset

Target

Revenue =1 if purchase, 0 otherwise
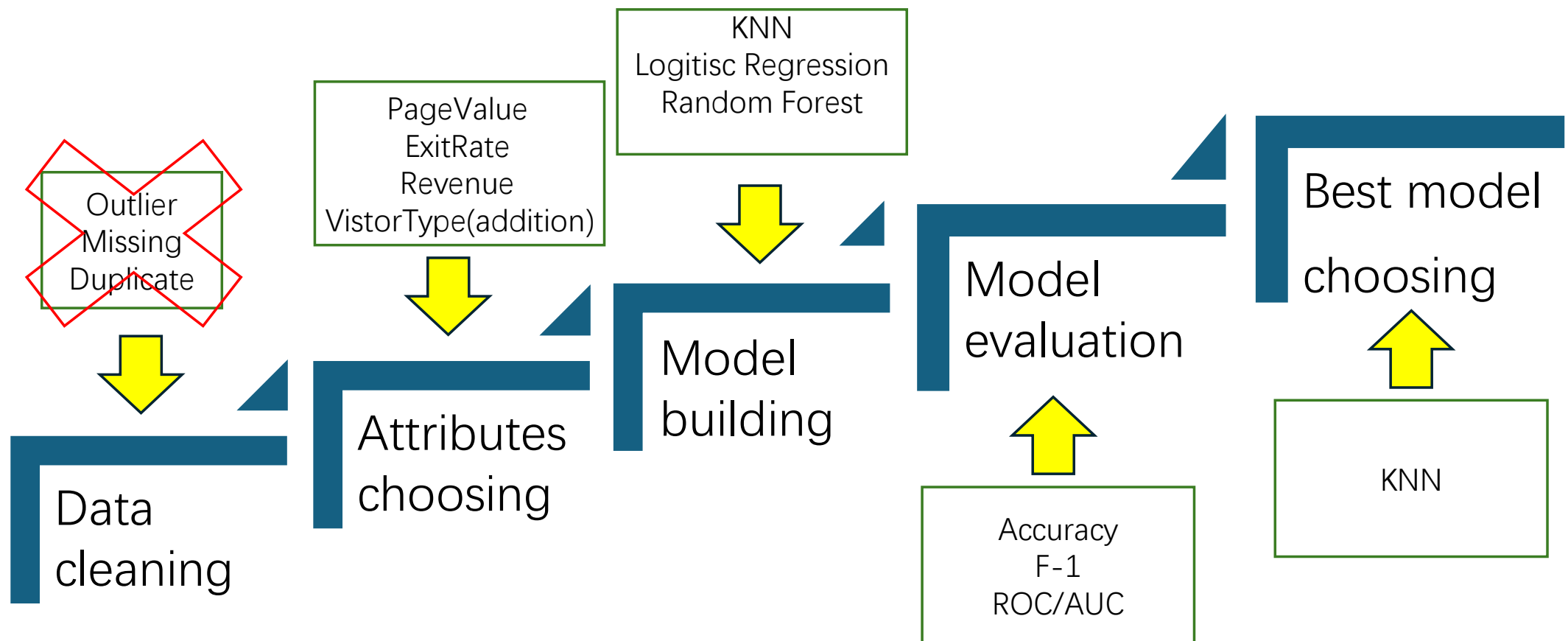
Data Quality

imbalanced (Revenue=1 ≈ 15%), clean numeric data

PS: Higher PageValue = user more likely to purchase;
Higher ExitRate = higher chance of losing customers.

Modeling Process

Outlier
Missing
Duplicate

Data
cleaning

PageValue
ExitRate
Revenue
VistorType(addition)

Attributes
choosing

KNN
Logitisc Regression
Random Forest

Model
building

Model
evaluation

Accuracy
F-1
ROC/AUC

Best model
choosing

KNN

Result

Relationship between PageValue, ExitRate, and Purchase Behavior

- **For data:**
  - PageValue and ExitRate can **effectively reflect** users' purchasing intentions
  - However, the dataset is **imbalanced**, leading to a lower F1-score for *Revenue=True*
- **For classifier:**
  - The KNN model's **accuracy is 88.4%** and **AUC is 0.8746**, indicating strong discriminative ability.
  - Despite this imbalance, **KNN provides the best trade-off between accuracy and recall** compared with Logistic Regression and Random Forest.
  - The model has **high recognition for non-purchase** and **medium recognition for purchase**
  - It is relatively **conservative** towards "potential buyers
- **This allows BOS to identify and prioritize high-value customers in some degree, improving conversion and ROI**
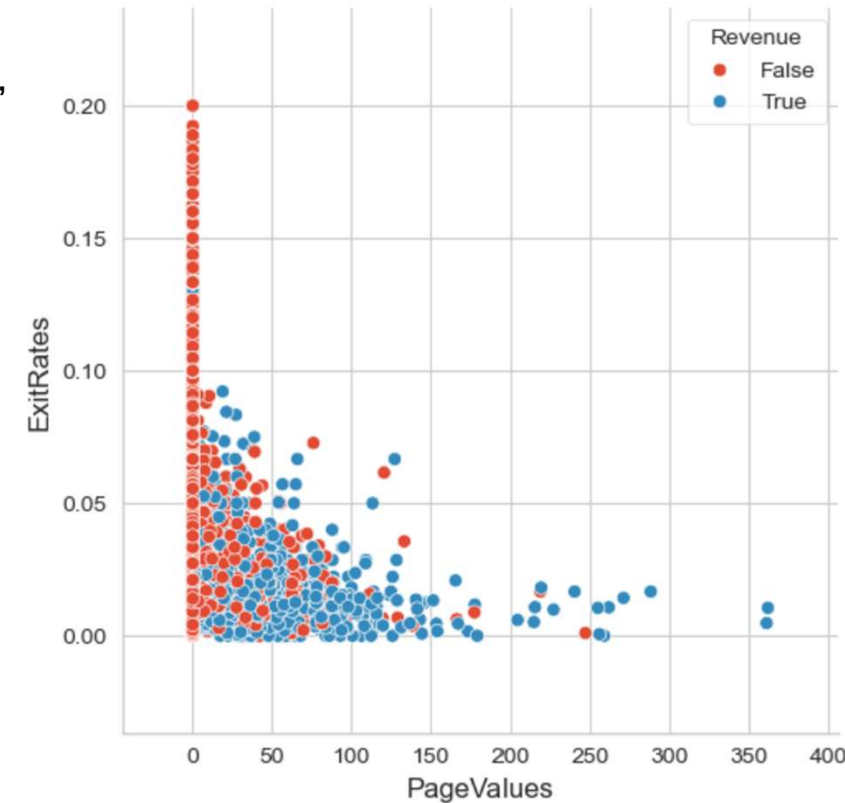
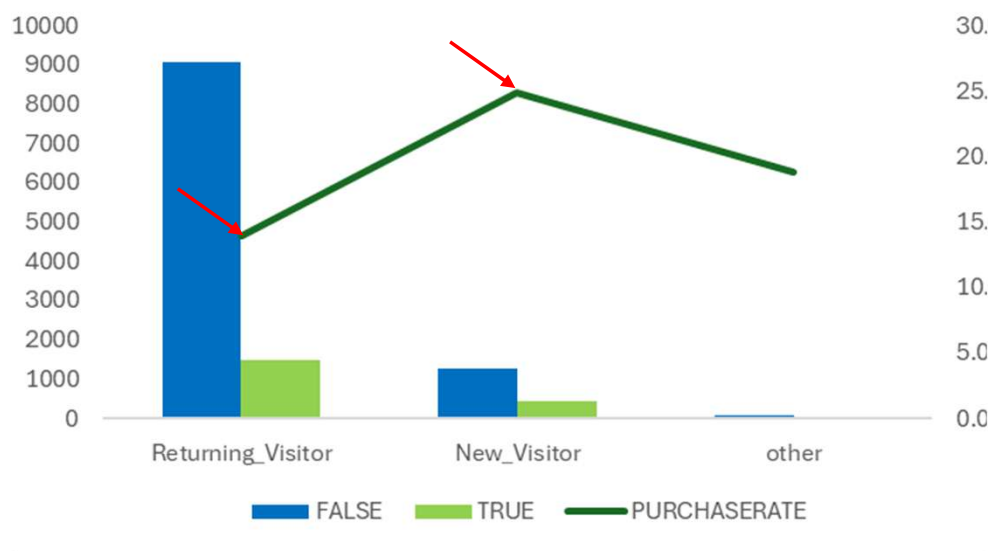Accuracy of our model is equal 88.4 %.

	precision	recall	f1-score	support
False	0.93	0.93	0.93	3128
True	0.63	0.61	0.62	571
accuracy			0.88	3699

Identifying High-Value Customers Based on Visitor Type

```
Accuracy of our model is equal 88.83 %.
               precision    recall  f1-score   support

      False       0.93       0.94      0.93      3128
       True       0.65       0.61      0.63       571

   accuracy                            0.89      3699
  macro avg       0.79       0.77      0.78      3699
weighted avg      0.89       0.89      0.89      3699
```

Knn's classification_report(add Vistor_Type)

relation between VISITORTYPE and Revenue

Returning visitor has largest volume but lowest revenue rate;
New visitors show higher purchase rate but lower volume.

- **Based on the previous analysis**, we added **VisitorType** into the original KNN model to better identify high-value customers.

- **For data:**
  - After including VisitorType as an additional feature, the model's **accuracy slightly improved** (from **88.4% → 88.83%**), though the dataset remains **imbalanced**

- **For classifier:**
  - KNN was maintained for consistency across comparisons
  - **Other could get same conclusion as last slide**
  - This shows that **VisitorType helps distinguish user behavior, which is helpful**

How to attract new visitor? How to improve the purchase rate of returning visitors?

Optimizing method for Visitor Type

Scatter plot to show relation between PV,ER,VT and Revenue
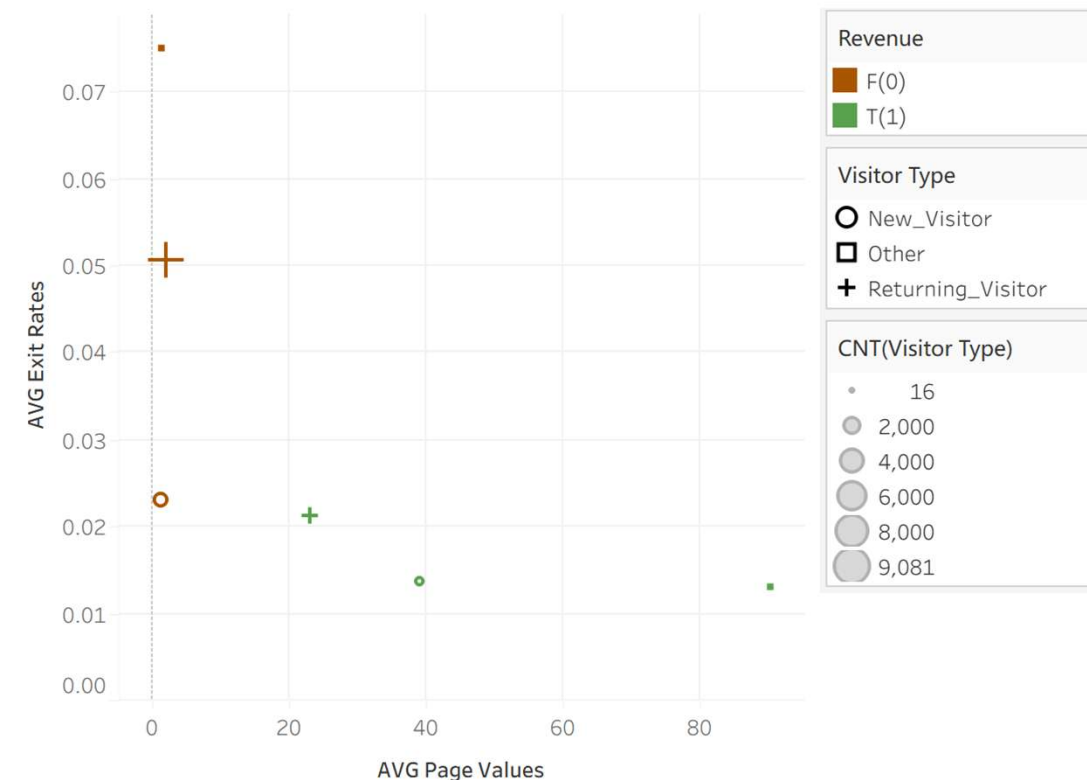
- The page design and recommendation logic should be optimized

- Advertisements!

- **New** Visitors convert best
  - offer first-purchase incentives
  - attractive web.

- **Returning** Visitors bring volume but low conversion
  - use retargeting or loyalty campaigns
  - given to promoting new product explanations, member privileges, and consumption rewards, etc

- **Other** has small sample size but high-Value
  - **Represent the potential high-net-worth client base**
  - Advertisements needs to be adjusted(address/form)

How Data and Cost Factors Influence the Model

Data:

imbalance, only 15% revenue=1

KNN model:

1.neighborhood voting the model is prone to predict the majority class (unpurchased)

2. **ignoring the real potential customers**

3.Need to update result time to time

method

Visitor type:

The behaviors of different types of visitors vary significantly. **It is best to discuss them in groups**

Cost:

1.The model is biased towards predicting False, **so less cost on web improved/ad publish**

2. **Adjust the weight to keep balance** between presion/recall

Adjustment

1.Balance the category ratio by using data resampling or adjusting the weight method.

2.Model by grouping visitors by type and optimize the prediction of new visitors and returning visitors respectively.

3.Is it precise targeting? It's better to spend more money to attract more people

Recommendation

The KNN model is highly accurate and easy to interpret. To enhance the model's ability to identify purchasers, the data can be reselected

- When the cost is limited, focus on Precision. Adjustments can be made based on business strategies or different sales cycles
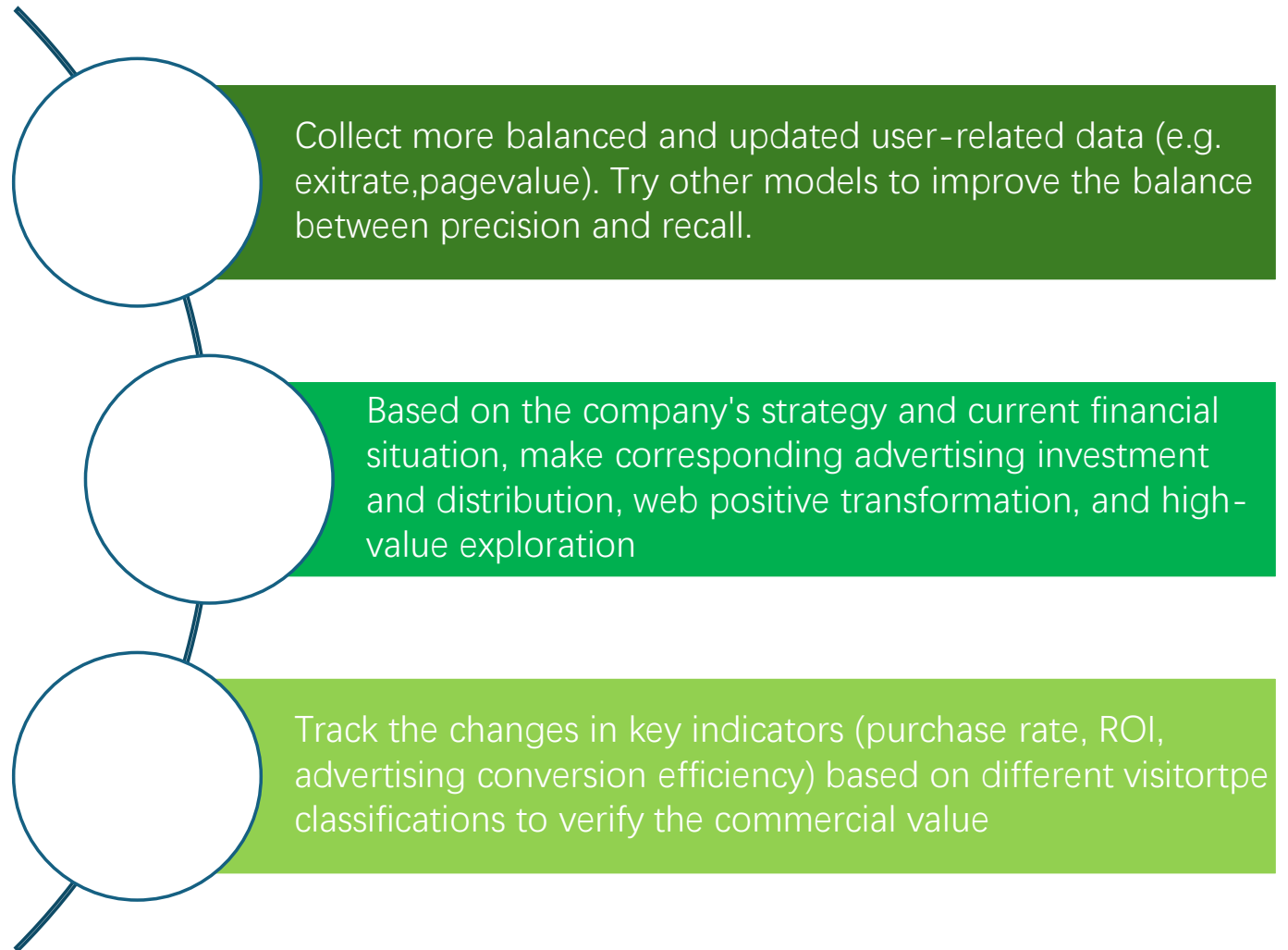
1.New visitor: First order discount, web design

2.Returning visits to users: Member benefits, remarketing activities

3.other: Advertising placement, web design

- retrain the model regularly to adapt to changes in user behavior and web page structure

Next Steps

Collect more balanced and updated user-related data (e.g. exitrate,pagevalue). Try other models to improve the balance between precision and recall.

Based on the company's strategy and current financial situation, make corresponding advertising investment and distribution, web positive transformation, and high-value exploration

Track the changes in key indicators (purchase rate, ROI, advertising conversion efficiency) based on different visitortpe classifications to verify the commercial value

Thank you so much

Appendix1: data, code and analysis

For dataset:TEB_V2_FOR_ASSIGNMENT_NO_ERRORS_online_shoppers_intention

[For more coding and analysis detail, please check:CSE5243_HW2_Template_V2 (1) (1)](#)

For visualization works on P8, please check here :[cse5243hw3 | Tableau Public](#)

Appendix2: adjustment

- Add VisitorType as an attribute into the original model for HW2

Appendix3: other resource

- P3 example: https://support.crazyegg.com/knowledge-base/case-study-ecommerce-website-increases-revenue-by-300-percent/

Appendix4: Reflections

- I realize that it is still difficult to actually apply specific data and analysis results to specific situations. In specific situations, it is necessary to consider the role of each attribute and the relationship between each factor. There are still many logics that are not particularly clear, and the methods that come up with might not be quite right either. Through this assignment, I have had a deeper understanding of the purpose of my work, learned about the relevant processes of making business analysis reports, and also mastered the application of office software