

# Project CLTV - Forecast customer's life time value

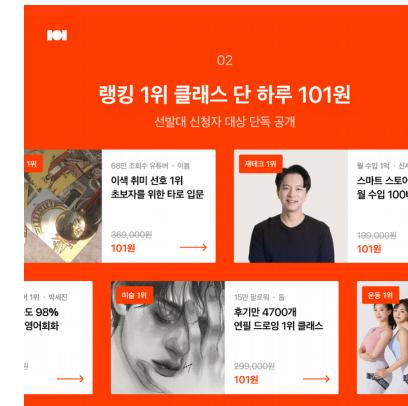
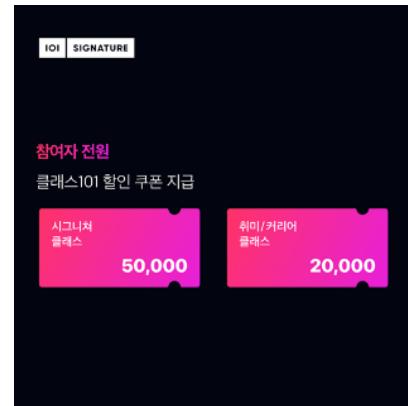
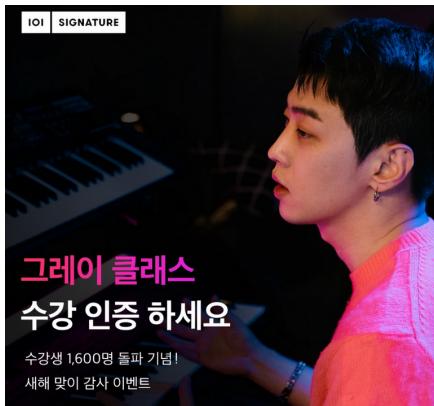
**Team member : HaLim Jun**

# Table of Contents

01. Background
02. Introduction of project
03. Model performance

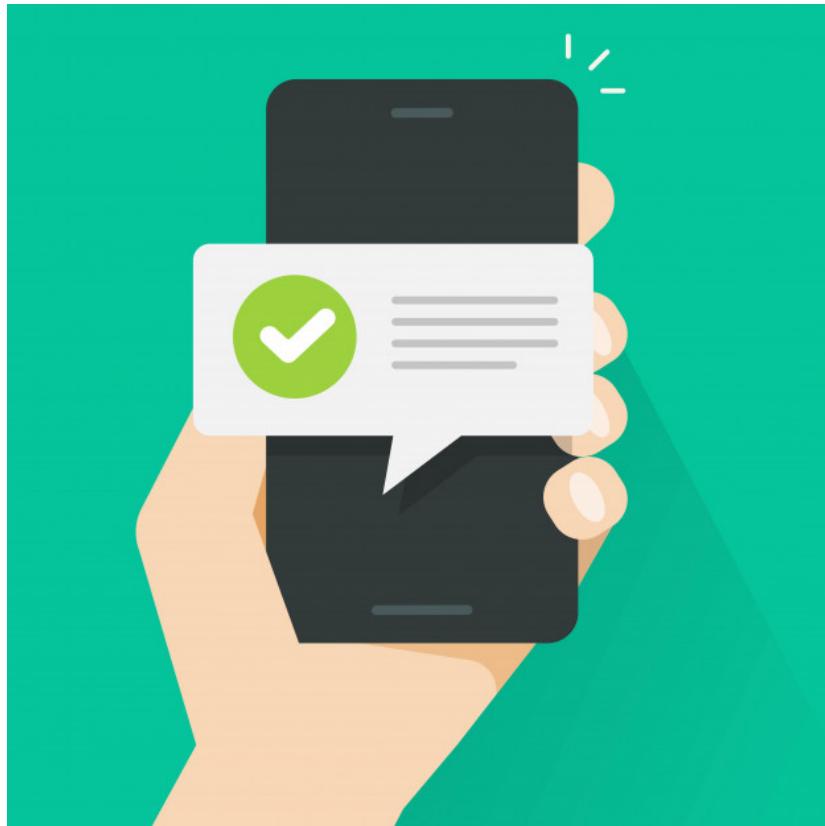
# 01 Background

What was the most attractive advertisement in recent days for you?



# 01 Background

How much discount should we give to which customers to maximize profit?



풀잎스쿨 역사상 최초 할인!  
온라인 풀잎 231,000원  
**165,000원!**

새해에는 안전하게  
온라인으로 열공하자!

- 풀잎 14기 정식모집 : 2020.12.14(월)  
- 진행 기간 : 2021.1.7 ~  
(11주\*2시간 or 8주\*3시간)

Deep learning course discount 66,000 krw!

A promotional graphic for a deep learning course. It features a blue-toned illustration of a person with blue hair, wearing headphones and holding a book titled '모두의'. The main text highlights a significant discount from 231,000 won to 165,000 won. It also mentions the start of registration for the 14th batch on December 14, 2020, and the duration of the course. A callout bubble on the right side provides a summary of the discount information.

# 01 Background

**How much discount should we give to which customers to maximize profit?**

If we can predict who is going to purchase how much,  
-> we will be able to optimize our marketing cost and  
maximize profit!

Expected revenue  
expected  
to decrease 50% for this  
group  
-> Increase level of  
discount!



## 02 Explanation

---

### 1. Target :

- Predict future transaction revenue of each customer using past transaction data
- Launch an ml ops model to be deployed in a business setting and develop a tool that marketers can easily utilize

2. Data : L-Point Data (2019/7 ~2019/9)

3. Phases :



## 02 Explanation 1) Model Set up

### Features

**Customer's transaction data during first two month  
(2019 July and August)**

- Transaction volume, number of product purchased
- Diversity in transaction
- Deviation in transaction volume
- Frequency and recency of transaction etc

### Target

**Customer's transaction data during first two month**

**Next month transaction volume prediction**

**2019/  
7~8**

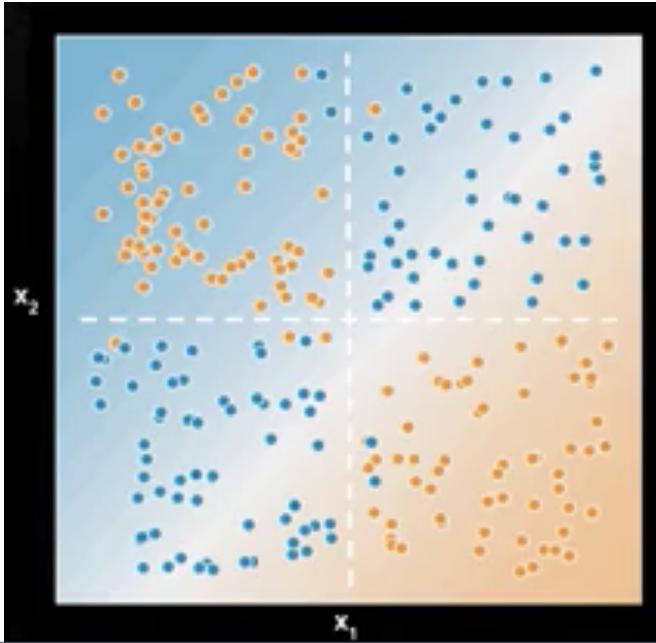
**2019/  
9**

## 02 Introduction 1) Model Set up

### Utilize feature cross to extract the relationship between feature variables

The probability for orange dot is related with  $x_1$  in a non linear way, thus linear model cannot learn such tendency

\*Can't draw a line that divides orange and blue line



$x_3$ , a product of  $x_1$  and  $x_2$ , is linearly related with  $y$ . The model's performance increase with such variable

\*Now it can be divided easily



$$x_3 = x_1 x_2$$

$$y = b + w_1 x_1 + w_2 x_2 + w_3 x_3$$

## 02 Introduction 1) Model Set up

### Model performance : log y Mean average error 0.61

Model: "sequential\_31"

Layer (type)	Output Shape	Param #
dense_152 (Dense)	(None, 4)	316
dense_153 (Dense)	(None, 8)	40
dense_154 (Dense)	(None, 8)	72
dense_155 (Dense)	(None, 8)	72
dense_156 (Dense)	(None, 1)	9

Total params: 509

Trainable params: 509

Non-trainable params: 0

**1. Total parameter : 509 (Light model)**

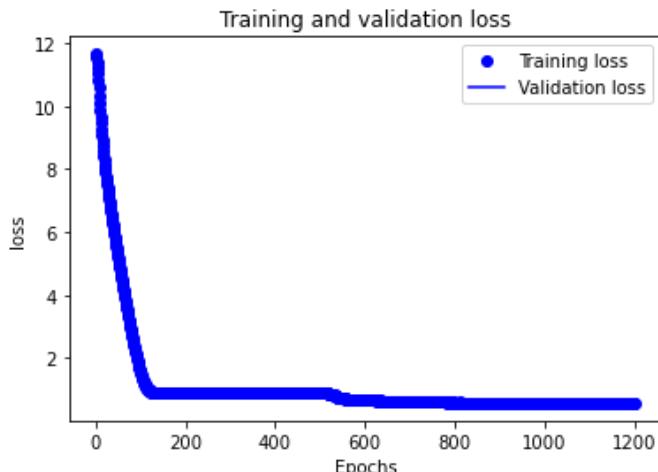
**2. Heavier model showed a worse performance due to overfitting**

- Tried to contain the number of parameters low in order to maintain the performance
- This is due to the large number of features put in

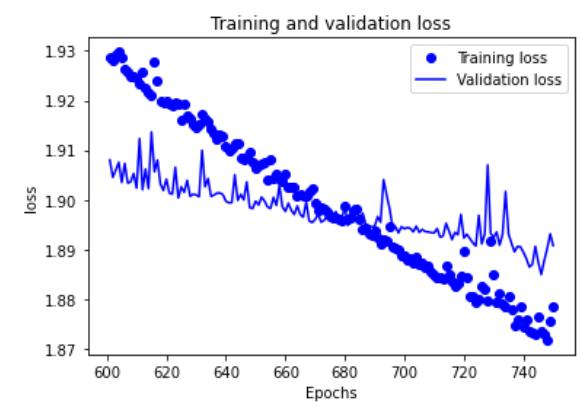
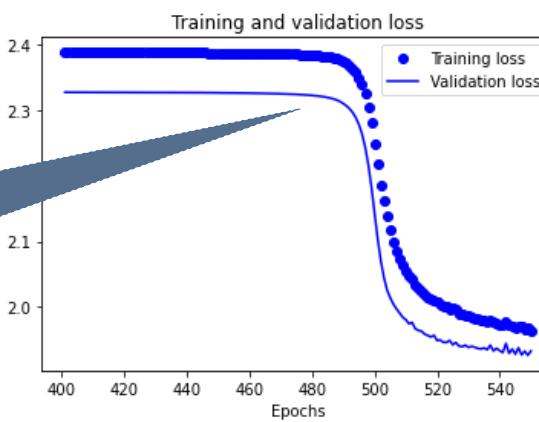
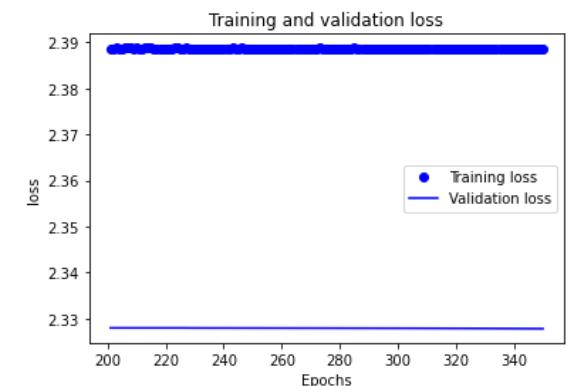
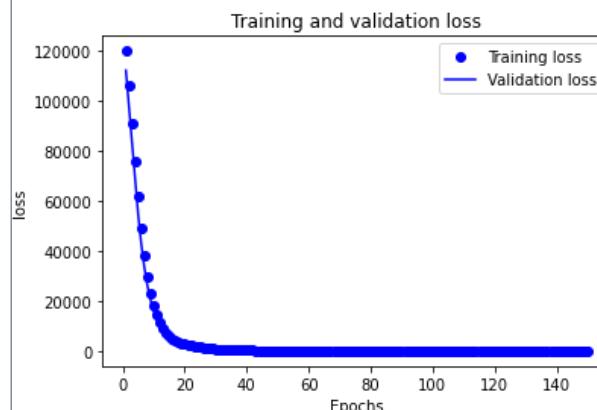
Minmax scaler Used, Batch size =512

## 02 Introduction 1) Model Set up

### Total Learning Curve



Rapid Learning → Saddle point → Learning again → Overfitting



Utilized Tanh to escape saddle point by speeding up the learning process

## 02 Introduction 2) Deployment

### Automation is a MUST to use machine learning in real business setting

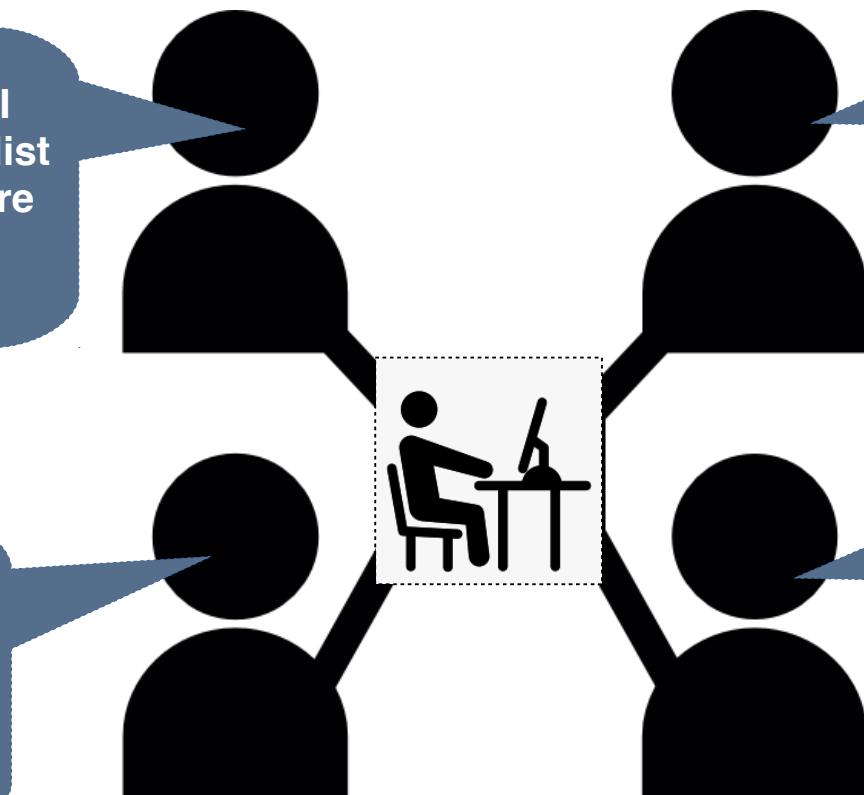
#### Scenario without Automation

Mr. Data Analyst, can I ask for the customer list who is likely have more than 30% decreased transaction?

Ms. Data Scientist, note that I haven't received the customers' predicted purchase value for next month

Mr. ML Engineer, may I have the list of customers who may purchase more than 1000 USD next month by the end of today?

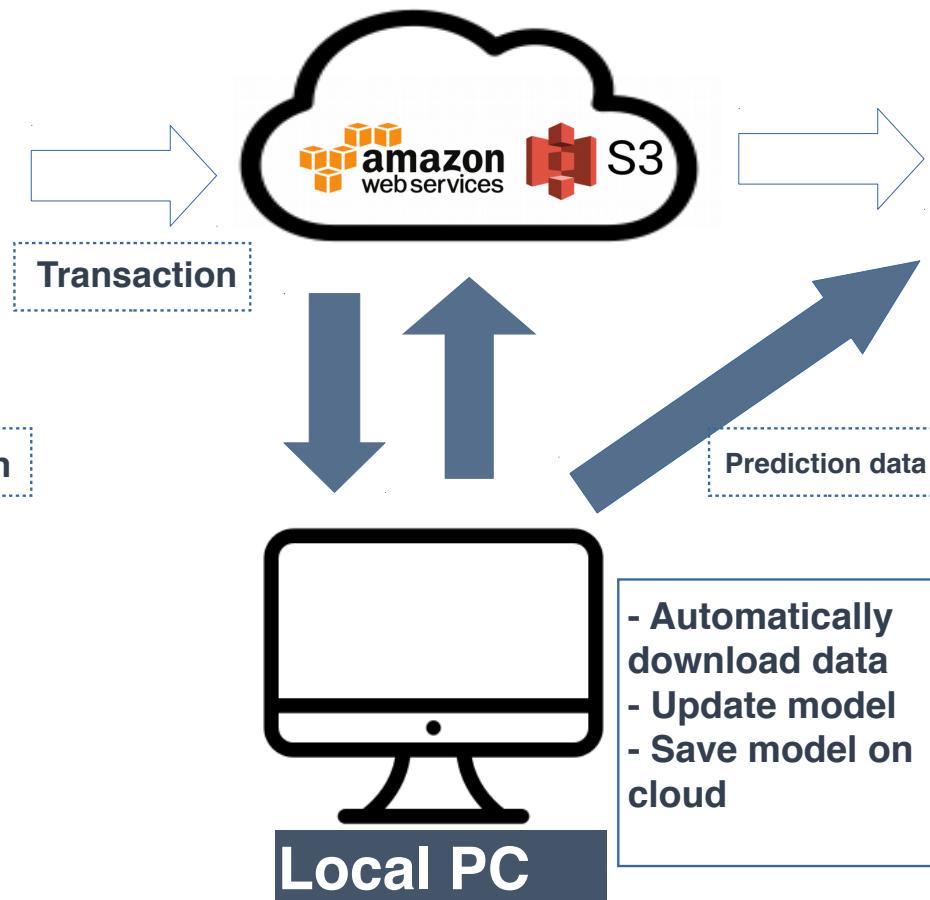
Mr. Data Engineer, Mr. Analyst is out of office. Could you get me the list of customers who is not going to purchase any next month?



## 02 Introduction 2) Deployment

Developed a semi automated ML Operation model using amazon S3

Transaction  
Data  
Update



Collaborative  
Dashboard tool



Easily download prediction insight and result



Marketer,  
Sales manager,  
CX manager

Developed

Not developed

hailey

**Amazon Simple Storage Service**  
**: Amazon Cloud Database**

**Access point overview**

Bucket projecttv	Network origin Internet	Region Asia Pacific (Seoul) ap-northeast-2
Amazon resource name (ARN) arn:aws:s3:ap-northeast-2:843258112828:accesspoint/hailey	Creation date December 17, 2020, 10:20 (UTC+09:00)	Access Bucket and objects not public

**Objects****Permissions**Drag and drop files and folders you want to upload here, or choose **Upload**.**Objects (3)**Objects are the fundamental entities stored in Amazon S3. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)
 List versions Delete Actions ▾ Create folder **Upload**

Find objects by prefix

&lt; 1 &gt;

<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	Transaction.csv	csv	December 20, 2020, 17:15 (UTC+09:00)	35.0 MB	Standard
<input type="checkbox"/>	weight_2020_12_20_23_54_44	-	December 20, 2020, 23:54 (UTC+09:00)	2.4 KB	Standard

**Amazon S3 Cloud Database Interface**

# 02 Introduction 2) Deployment

**Set up automatic deployment process : Download data from Amazon S3 Cloud  
-> Machine learning on Local PC -> Upload model weight**

Used Boto3 (A python package that enables interaction between python & AWS)

## Data download

```
def download_data(path=''):
    now=datetime.datetime.now()
    date_time = now.strftime("%Y %m %d %H %M %S")
    file_name=date_time+'source_file.csv'
    s3 = boto3.client('s3', aws_access_key_id='AKIAJFG4KPBJCLYUUCIA',
    aws_secret_access_key='gRSykBpJITY[REDACTED]MR')
    s3.download_file('projectltv', 'Transaction.csv', file_name)
    df=pd.read_csv(path+file_name)
    return df
```

## Upload model weight

```
def upload_model_weight(weight_path,bucketname):
    s3 = boto3.client('s3', aws_access_key_id='AKIAJFG4KPBJCLYUUCIA',
    aws_secret_access_key='gRSykBpJI[REDACTED]')
    saved_model=load_model(weight_path, compile=False)
    model_to_json=saved_model.to_json()
    with open('model_to_json.json', "w") as json_file:
        json_file.write(model_to_json)
    now=datetime.datetime.now()
    date_time = now.strftime("%Y %m %d %H %M %S")
    as_file_name='weight_'+date_time
    s3.upload_file('model_to_json.json', bucketname, as_file_name)
    return as_file_name
```

Access AWS  
cloud with AWS  
secret key

The model weight files  
should be converted into  
Json format in order to be  
reused

## 02 Introduction 2) Deployment

---

### Demo Video

<https://youtu.be/EdNHKsAGVos>

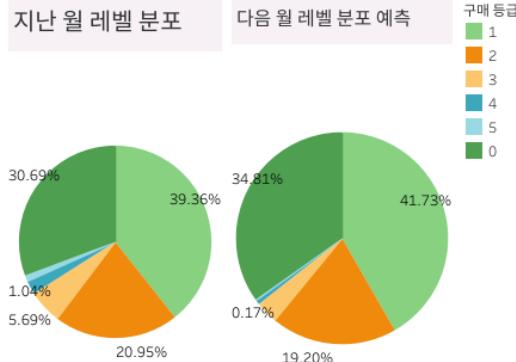


## 02 Introduction 3) Collaborative Dashboard

### A customer marketing forecast dashboard built on Tableau site

Sales 예측 대시보드

지난 월 레벨 분포



다음 월 레벨 분포 예측

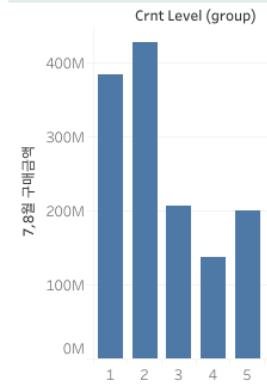
구매등급

1  
2  
3  
4  
5  
0

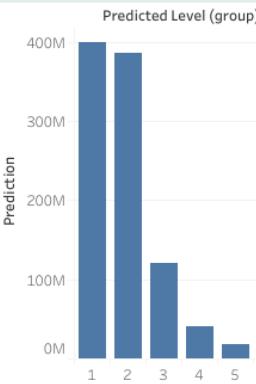
다음월 레벨 간 이동 예측(절대수치)

Crnt Level (..)	Predicted Level (group)					
	1	2	3	4	5	0
1	1,417	220	9	1	592	
2	476	646	42	2	79	
3	51	180	93	3	11	
4	22	36	45	22	10	
5	4	18	13	12	5	
0	510	41	1		1,272	

레벨별 매출(전 2개월 평균)



레벨 별 매출 예측(다음월)



다음월 레벨 간 이동 예측(상대수치)

Crnt Level (..)	Predicted Level (group)					
	1	2	3	4	5	0
1	60.58%	9.41%	0.38%	0.04%	29.59%	
2	38.23%	51.89%	3.37%	0.16%	6.35%	
3	15.09%	53.25%	27.51%	0.89%	3.25%	
4	16.30%	26.67%	33.33%	16.30%	7.41%	
5	6.45%	23.03%	20.97%	19.35%	16.13%	8.06%
0	27.96%	2.25%	0.05%		69.74%	

The dashboard was created to facilitate data sharing within business setting

Anyone without SQL knowledge can easily access the dashboard through the internet

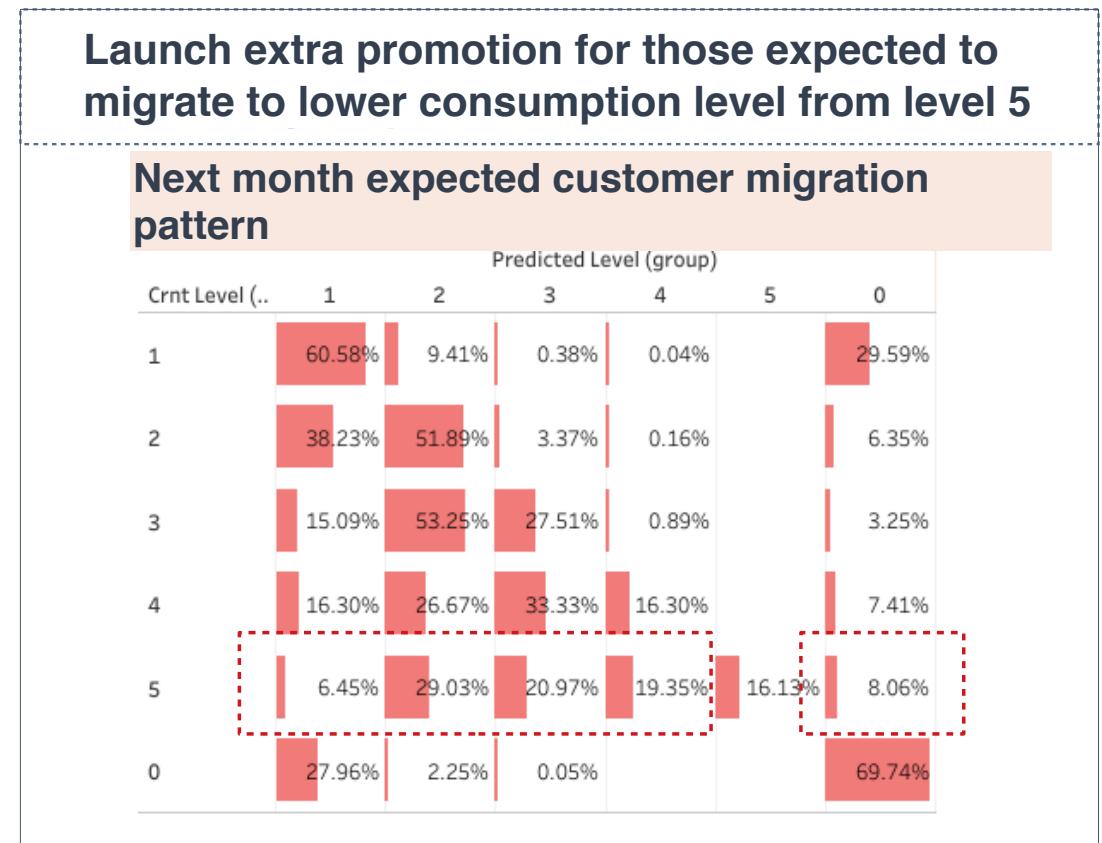
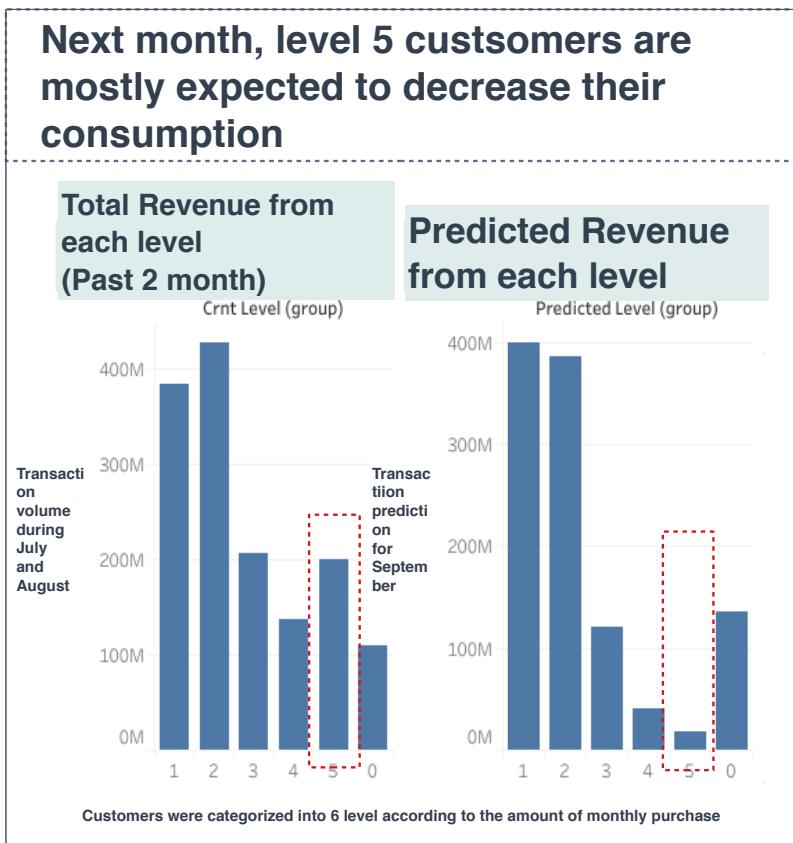
- With simple signing up on tableau site, can access the dashboard through web
- We can download relevant data by just clicking on the graph

대시보드 URL : <https://prod-apnortheast-a.online.tableau.com/t/haileyproject/authoring/project1/Dashboard1#2>

구매 금액에 따라 0 레벨 ~ 5 레벨로 구분 10 만원 이하 , 10 만원 ~25 만원 , 25~50 만원 , 50~85 만원 , 85~150 만원 , 150 만원 이상으로 구분

## 02 Introduction 3) Collaborative Dashboard

### Example case) Using the dashboard for preemptive marketing for revenue decrease



dashboard URL : <https://prod-apnortheast-a.online.tableau.com/t/haileyproject/authoring/project1/Dashboard1#2>

구매 금액에 따라 0 레벨 ~ 5 레벨로 구분 10 만원 이하 , 10 만원 ~25 만원 , 25~50 만원 , 50~85 만원 , 85~150 만원 , 150 만원 이상으로 구분

## 02 Introduction 3) Collaborative Dashboard

---

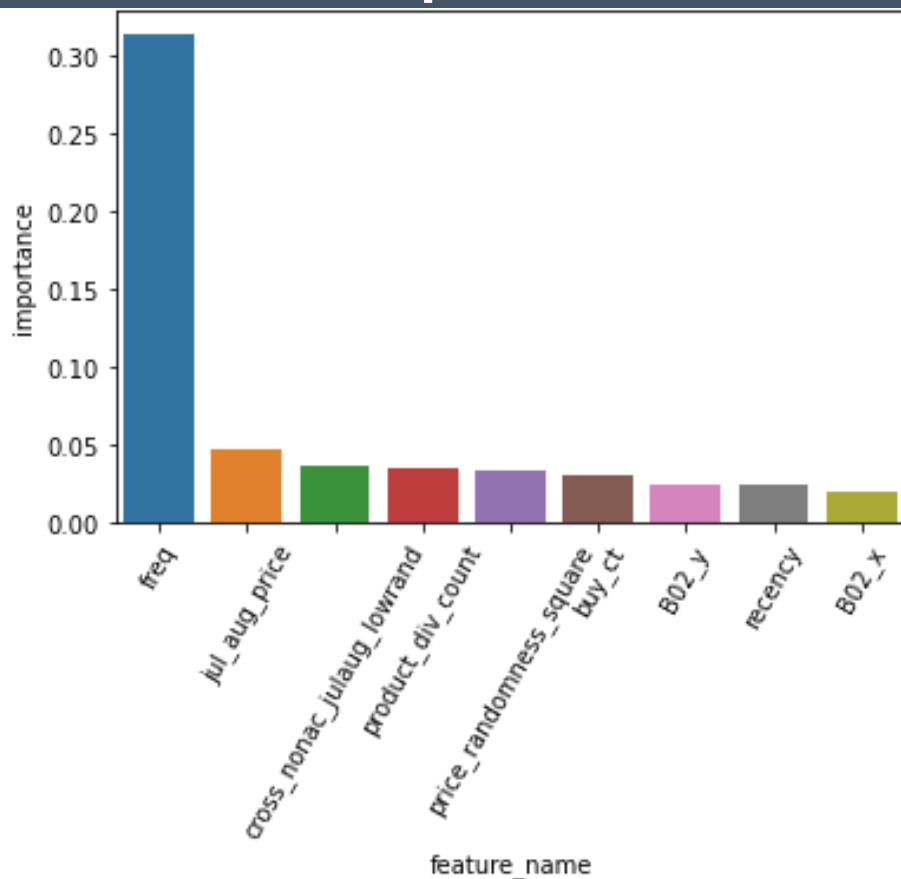
**Just with easy clicks, anyone access the model's insight and outcome without knowledge of SQL**

<https://youtu.be/dLwV3EAsZc4>

**demo Video**

## 03 Model Performance

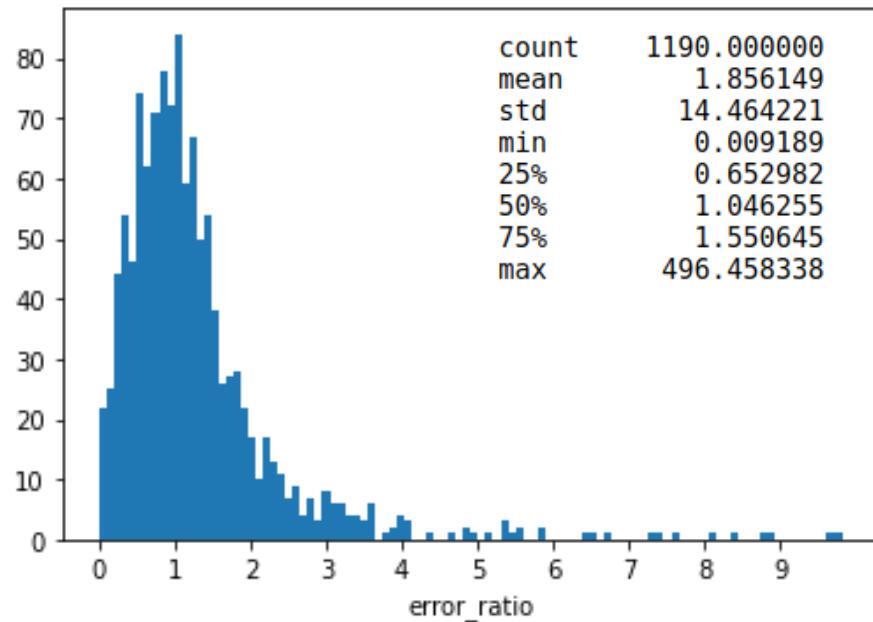
Transaction frequency, total transaction volume during first two months, randomness in transaction volume, customer recency turned out to be important feature for prediction



# 03

## Model Performance

**Model performance : Mean average error 0.61 for log of y (y = transaction amount y). Around 60% of customers were predicted with 0.62 ~1.5 accuracy**



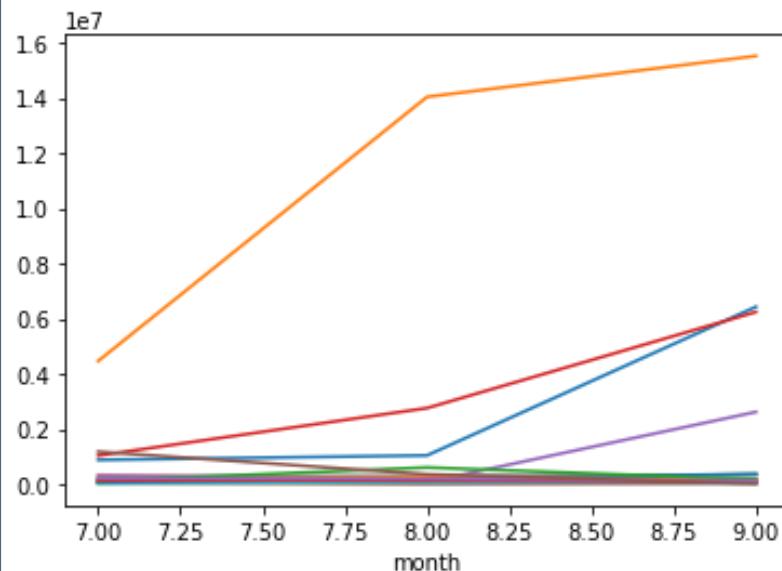
**Error ratio = Real value of transaction/Predicted value of transaction**

# 03

## Model Performance - diagnosis

When there is sudden and drastic change in behavior (eg, transaction volume or quantity) model shows weaker performance

### Most under-valued customer's trend in transaction volume



Sudden upsurge of transactions are not well forecasted  
-> Need more data to overcome such problems  
(eg, Behavioral log data online, promotions, longer term data)