Meal Delivery Platform’s On-Time Performance Improvement

——Based on a meal delivery dataset in Shanghai

DMS3003 Group 7 Final Report

116020139 Liang Yuming 116020356 Zhi Zhaoyan

118020041 Lin Yicheng 116020276 Xia Yingying

Catalog:

Abstract

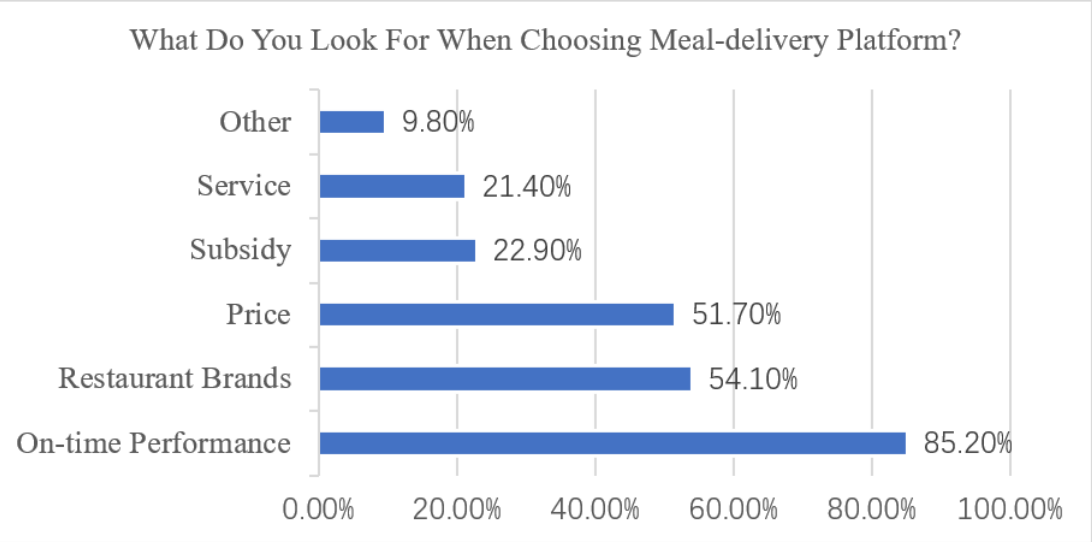
In this article, we used a dataset from a meal delivery platform in Shanghai to analyze some problems related to meal delivery platforms' on-time performance (whether the meals are delivered within the time limits set by platform). We first analyzed the relationship between customers repurchase rate and on-time performance of meal delivery platform. We then studied the relationship between on-time performance and some other variables related with meal delivery. Goodness of fit, multicollinearity test, and endogeneity test are conducted to make sure the result is reliable.

After our analyses, we conclude that on-time performance indeed affects customers repurchase, and thus impact platform companies' revenues. Among all the variables, riders' waiting time in the restaurant is the most significant factor that influence on-time delivery performance. Since riders' waiting time is dependent on restaurants' preparation time, we make suggestions focusing on improvements of restaurants' preparation speed. Also, restaurants which have high preparation time also have higher variance in terms of on-time performance. We therefore suggested that platforms should make use of recommendation system to improve restaurants' preparation time performance.

Statement of Needs

In recent years, online meal delivery platforms have grown rapidly in China with Meituan Waimai and Eleme dominating the market. However, as the market develops and subsidy level decreases, customers care more and more about service quality of the platform. Among all the issues that may affect service quality, customers particularly focus on on-time delivery performance.

When customers order meals from meal delivery platforms, platforms always give an estimated arrival time to customers in order to help customers plan their schedules. However, the estimation is not 100% accurate. In some cases, riders may fail to deliver meals on time. Under this scenario, customers may feel unhappy and they are not willing to choose this platform again. According to a survey conducted by analysys.com, 85.2% of customers care about on-time performance when choosing a meal delivery platform, ranking the 1st among six factors (Yang, 2015).[1]



Therefore, it is very important to improve on-time performance in order to increase return purchases, which means more revenues brought to the platform. Those impacts should be proved and quantified to help platforms make better decisions.

Although meal delivery is a traditional service for restaurants, LBS (Location Based Service) type meal delivery is actually an emerging direction. Compared with traditional model, the new model doesn't rely on "call-cook-deliver" chain, but rather an application-based resource management system. It was built because of the higher coverage of GPS, smartphones, and 4th generation network. Traditional studies concerning meal delivery were more or less focusing on route optimization and order distribution towards riders, where there are hardly any studies concerning service level improvements. What makes our project stands out is that we set customers' satisfactions as a start point and end goals, and we use data analytic techniques to improve on-time performance to gain more revenues.

Research Background

a. Research Questions

This research focuses on providing a solution to address delivery overtime issue to improve customer satisfactions. In this way, two related research questions have been raised.

Firstly, how does the on-time performance affect customers' satisfactions? Does the on-time performance have a significant impact on customers' satisfactions, and thus influence revenue of the platform? Secondly, if the answer is yes to the first question, what factors contribute most to the on-time performance?

To answer these two research questions, one important definition will be used across the paper. This paper used relative deliver time proportion (RDTP for short) to represent on-time performance. We define the Relative Delivery Time Ratio (RDTP, for short) as: (finished time - placed time) / (required time - placed time). As the time that delivery orders last can vary dramatically among different orders, we use relative ratio of this two rather than the absolute difference to standardize the impact of overtime on customers. When the order delivery is overtime (when RDTP > 1), we consider the on-time performance with such RDTP as inefficient performance.

b. Premises

To delve into this question, several premises are made:

The required time set by the company is assumed to be reasonable. If the delivery arrives at the required time, customer satisfaction is high. Such assumption is needed because the required time works as the benchmark in the definition of overtime delivery. If the required time is not reasonable, then it is meaningless to discuss about overtime delivery phenomenon. Moreover, even there might be some inefficiencies regarding platforms' delivery arrival time prediction algorithm, it is hardly possible and also too costly for the author to provide managerial suggestions to improve the efficiency on required time in this paper. It is mainly because the external researcher cannot get access to the specific design of the prediction algorithm, making it impossible to provide improvements. Additionally, improving the algorithm could be a very costly managerial decision in business world. Therefore, to truly improve the customer satisfaction level, this paper assumes the required time is reasonable and focuses on other factors influencing RDTP.

Riders' working time is equivalent each day. This can be justified by the fact that the riders of this company were all full-time workers back to 2015. Such assumption is used in hypothesis to justify the indicator for rider’s efficiency.

c. Dataset

The dataset is an excerpt of Meituan’s meal delivery data in Shanghai from Sep. 1st, 2015 to Sep. 25th, 2015, in which 19 days are workdays and 6 are weekends.

To avoid biases caused by potential incorrect data, the original dataset was cleaned such that data with missing values were skipped, and noises and outliers were detected and deleted.

The original dataset has 133,769 rows (order entries) and 14 attributes. While the preprocessed dataset has 70,547 entries. There are 1,060 riders, and on average each rider delivers 5.05 orders per day. There are 3,012 restaurants, and on average each restaurant has 1.78 orders per day. The average order price is 70.75 RMB and riders can get on average 6.57 RMB per order. Among all the orders, 76.48 percent were able to deliver on-time.

d. Hypothesis

Hypothesis 1:

This paper raised first hypothesis: Increase in repurchase time will not lead to decrease in per order transaction. This hypothesis is important to test so we can make sure that the impact of increasing repurchase time can lead to an increase of revenue, given the number of consumers is stable.

Hypothesis 2:

To answer first research question investigating the relationship between on-time performance and customer satisfaction level, we first need to define and evaluate customer satisfaction level. In 2015, the platform didn’t have such function as customer's rating, thus this paper uses repurchase rate as a way to indicate customer stickiness and customers' satisfaction level, as satisfied customers will be more likely to make orders next time. This way, this paper will compare the times customers make return purchase to evaluate its satisfaction level.

As stated above, more than 85% of customers in 2015 (our dataset was recorded at the same period of time) customers care about on-time performance when choosing a meal delivery platform, and the preference still exists nowadays. Therefore, to answer first research question, hypothesis 1 is defined as follows: better on-time performance leads to higher customer satisfaction.

Hypothesis 3:

To delve into second research question which investigates the factors influencing the on-time performance (RDTP), we will break it down to 6 hypotheses

Hypothesis 3a:

Riders’ waiting time proportion (WTP for short), which evaluates the time the riders have to wait for restaurants over the total delivery time, may impact the RDTP.

When a rider arrives at a restaurant, the time the rider waits can be uncertain and long, thus influences the overall RDTP and also leads to inefficiency for the time riders waste. We define WTP as: (leave time - arrive time) / (require time - place time). A high WTP matches with the scenario that orders have exceed restaurants’ cooking capacity, and riders have to wait for a long time till the cooking is done.

We conjecture that if the cooking time occupies a larger part of the whole delivery time (defined as: require time-place time), the RDTP will be larger (larger RDTP can be good or bad. It depends on how much time left for riders' deliveries).

Hypothesis 3b:

This paper will then look at riders' influence on the RDTP.

We assumed riders' experience level will negatively correlated with RDTP, as the more experienced a rider is, the less likely he or she will delay the order.

From the data provider, we learnt that all the riders in 2015 were full-time riders, therefore, we defined level of experience through the number of working days of the particular rider within dataset's coverage period.

Hypothesis 3c:

Furthermore, we conjecture that the riders' efficiency (defined as: average number of finished orders per working day of certain rider) will impact his or her on-time performance (defined as: average RDTP of certain rider). The more efficient the rider is, the less likely it will delay, so our hypothesis is that riders' efficiency is negatively correlated with the RDTP.

Hypothesis 3d:

We conjecture that the customers' order time (place time) may impact RDTP. We speculate that the RDTP may be higher during peak meal time 11:00-13:00 and 17:00-19:00. Since during the peak time, there are more chances of larger quantities of orders for the platforms. Larger quantities of orders might lead to over capacity of the system, leading more chances of late delivery. Moreover, during the peak time, there is higher chances of traffic jam due to more mobility of people and similarly large number of orders for the competitors.

In this way, we conjecture customers' order time during peak time might lead to larger RDTP.

Hypothesis 3e:

The weather condition could be important as well. Bad weathers will lead to more uncertainty in the riders' delivery process, thus leading to higher RDTP. We define rainy day as bad weather, non-rainy day (including sunny day and windy day) as nonbad weather. The data related to weather condition is gained through Tianqihoubao open data platform.

Therefore, we conjecture bad weather condition might lead to higher RDTP.

Hypothesis 3f:

Lastly, days in a week can also affect RDTP. In the weekdays, there could be chances people (especially for working people and white-collar workers) are busy so they consider convenience as the higher priority, therefore during the weekdays, people are likely to place more orders than those in weekend. With more orders, there are chances of overload or uncertainties. Therefore, we conjecture weekdays might lead to higher RDTP.

e. Methodology

We first calculated the correlation coefficient and applied a t test to study the relationship between repurchase time and per order transaction (for hypothesis 1). Then we use F-test to study the relationship between repurchase behavior and RDTP (for hypothesis 2). Finally, we build a linear regression model to study the significant factors that related with RDTP (for hypothesis 3). (summary of method see Table 1)

Method Table [Table 1]

Model	Question	Hypothesis	Method
1	Does increase in repurchase time decrease per order transaction	1	Correlation Coefficient
2	Does RDTP affect repurchase time	2	T-test of Correlation Coefficient
3	What are important factors affecting RDTP	3	Linear Regression model

Model Analysis and Result

Model 1: Correlation coefficient of repurchase time and per order transaction

We calculate the correlation coefficient（function see figure 1）between repurchase time and per order (variable definition see Table 2). The result is -0.068 (see Appendix 1). Since the correlation coefficient is less than 0.1, we can conclude that the correlation between per order transaction and repurchase time is small. When number of consumers is stable, increasing repurchase time can increase revenue for both restaurants and platform 【revenue=number of consumer * (repurchase time per consumer+1) * per order transaction】.

Correlation coefficient function between repurchase time and per order transaction [Figure 1]

$$r = \frac{Cov(Repurchase\ Time,\ \ Per\ Order\ Transaction\ Value)}{\sigma_{Repurchase\ Time}\sigma_{Per\ Order\ Transaction\ Value}}$$

Variable Definition [Table 2]

Variable	Definition
Repurchase Time	number of order one consumer make during the dataset's period
Per Order Transaction	average order price for one consumer

Model 2: T-test of the correlation coefficient between repurchase time and RDTP

We calculate the correlation coefficient（function see figure 2） between repurchase time and RDTP (variable definition see Table 3). The result is -0.153 with p-value at 0.14426 (see Appendix 2) in T-test (t-test function see Figure 3). Since the p-value is less than 0.15, we can reject the null hypothesis (correlation coefficient is zero) at 15% significance level. Therefore, we can conclude that there is negative correlation between RDTP and repurchase time. When RDTP increase, repurchase time decrease.

Correlation coefficient function between repurchase time and RDTP [Figure 3]

$$r = \frac{Cov(Repurchase\ Time,\ \ RDTP)}{\sigma_{Repurchase\ Time}\sigma_{RDTP}}$$

Variable Definition [Table 3]

Variable	Definition
Repurchase Time	number of order one consumer make during the dataset's period
Average RDTP	average RDTP for consumers under the same repurchase time note: RDTP=(finish time-place time)/(required time-place time)

T-test for Correlation Coefficient [Figure 3]

n represents number of sample (range of repurchase time is 0-49, so n is 50), r represents correlation coefficient.

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

Model 3: Linear Regression Model for RDTP

We use Bidirectional elimination method to regress RDTP on several variables. (Model see figure 4, variable definition see figure 5).

Linear Regression Model Variable [figure 4]

$$RDTP = \alpha + \sum^{14} X_i$$

X_1 : $rider\ efficiency$ ($average\ number\ of\ finished\ orders\ per\ day\ of\ certain\ rider$)

X_2 : $working\ days$ ($total\ working\ days\ of\ rider\ during$ 25 $days$)

X_3 : $rider\ waiting\ time\ proportion$ $\dfrac{leave\ time - arrive\ time}{required\ time - place\ time}$

X_4 : $lunch\ peak$ ($order\ time\ is\ during$ 11 $a.m.$ to 1 $p.m.$: 1; $otherwise$: 0)

X_5 : $dinner\ peak$ ($order\ time\ during$ 5 $p.m.$ to 7 $p.m.$: 1; $otherwise$: 0)

$X_6 - X_8$: $dummy\ variable\ for\ shower\ rain$; $middle\ rain$; $heavy\ rain$;

$X_9 - X_{14}$: $dummy\ variable\ for\ working\ day$, $from\ sunday\ to\ Friday$

The regression model yields the following result, shown in figure 5:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   RDTP   R-squared:                       0.544
Model:                            OLS   Adj. R-squared:                  0.544
Method:                 Least Squares   F-statistic:                     5988.
Date:                Sun, 10 May 2020   Prob (F-statistic):               0.00
Time:                        02:24:06   Log-Likelihood:                 6309.0
No. Observations:               70261   AIC:                        -1.259e+04
Df Residuals:                   70246   BIC:                        -1.245e+04
Df Model:                          14
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.5502      0.005    117.640      0.000       0.541       0.559
x1            -0.0065      0.000    -18.246      0.000      -0.007      -0.006
x2            -0.0012      0.000     -7.582      0.000      -0.001      -0.001
x3             0.7427      0.003    266.802      0.000       0.737       0.748
x4             0.0798      0.002     40.515      0.000       0.076       0.084
x5             0.0079      0.002      3.551      0.000       0.004       0.012
x6            -0.0026      0.003     -0.996      0.319      -0.008       0.003
x7            -0.0319      0.005     -6.388      0.000      -0.042      -0.022
x8             0.0020      0.003      0.683      0.495      -0.004       0.008
x9             0.0379      0.003     11.063      0.000       0.031       0.045
x10            0.0117      0.004      3.254      0.001       0.005       0.019
x11            0.0115      0.003      3.388      0.001       0.005       0.018
x12            0.0020      0.003      0.591      0.555      -0.005       0.009
x13            0.0139      0.004      3.455      0.001       0.006       0.022
x14            0.0080      0.003      2.306      0.021       0.001       0.015
==============================================================================
Omnibus:                    80726.457   Durbin-Watson:                   1.730
Prob(Omnibus):                  0.000   Jarque-Bera (JB):         68637077.027
Skew:                           5.240   Prob(JB):                         0.00
Kurtosis:                     155.760   Cond. No.                         177.
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

In figure 4, x1 represents rider efficiency, x2 represents rider experience, and x3 represents rider's waiting time proportion. x4-x5 are the dummy variables of peak time, and the condition of non-peak time is ignored. x6-x8 are the dummy variables for weather condition, and the condition of non-rainy day is ignored. X9-x14 are the dummy variables for day of week, and the condition of Friday is ignored.

As we can observe from figure 5, the adjusted R2 for the model is 0.544, which indicates that our model can partly explain for the on-time performance of meal delivery. The coefficient for variables of weather condition and day of week are insignificant, which rejects our hypotheses 3e and 3f. Rider's waiting time in the restaurant is the most significant factor, and it confirms our hypothesis 3a that overtime delivery is more likely to happen when the rider is stuck at the restaurant waiting till the cooking is done. The coefficients for rider efficiency and rider experience are both negative with significance, which confirms hypothesis 3b and 3c that more efficient and experienced rider is less likely to deliver order overtime. The coefficient for peak time is positive with significance, indicating that delivery required to be delivered during peak time are more likely to be delivered overtime, which verified hypothesis 3d.

Model Adjustment and Limitation

We improved interpretation of the linear regression model by testing multicollinearity, heteroscedasticity and endogeneity problems and make corresponding adjustments.

Multicollinearity

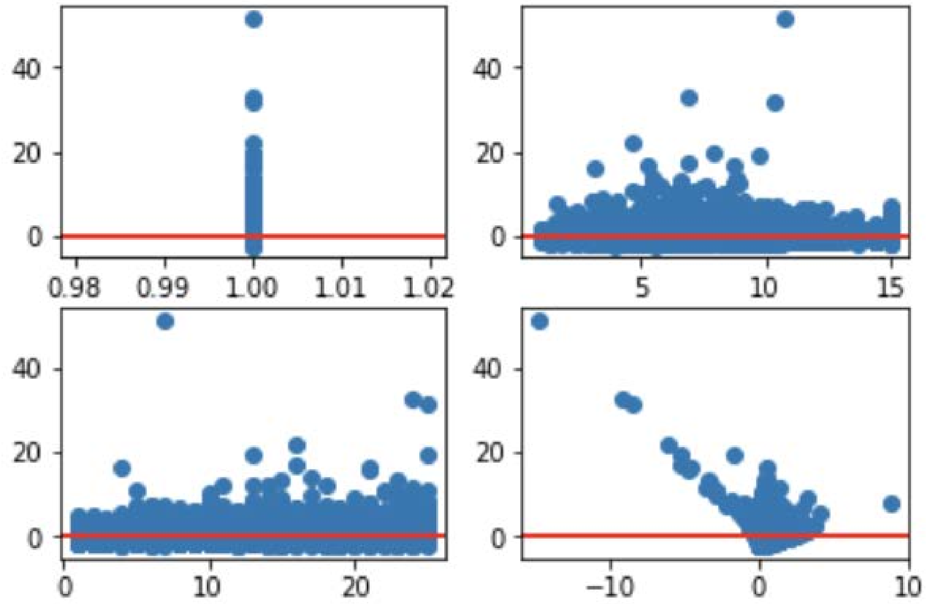
Multicollinearity can lead to higher variance of model coefficients and reduce the significance level. We tested the correlation among different attributes. Specifically, rider efficiency may intuitively correlate with rider's experience level, which is defined ad working days over the whole period of time, so we test the correlation coefficient and get a result of 0.164 (see appendix 3). It is less than 0, hence there is no severe multicollinearity in the model.

Heteroscedasticity

We first plot the distribution of error term on the 4 variables separately (top left: constant term; top right: rider efficiency; bottom left: rider experience level; bottom right: rider's waiting time proportion) (see figure 6), the fluctuating distribution indicates that there is heteroscedasticity. Then we run a white-test and the result shows heteroscedasticity (see appendix 4). Since heteroscedasticity may increase the variance of coefficient and reduce precision, this can be one limitation of our model.

Error term distribution on different variable [Figure 6]

Y dimension: error term; X dimension: variable



Endogeneity

Since it is common for a linear regression model to have endogeneity and many variables related to RDTP are not considered in this model, we tend to reduce the endogeneity. Because waiting time proportion (written as WTP in the following) is significant in explaining RDTP with much higher coefficient and larger t-value than other variables, we tend to find instrument variables for WTP. However, we fail to find proper IVs which have high correlation with WTP (several trails can be found in appendix 5). Therefore, the biasness brought by endogeneity is one limitation of the model.

In conclusion, the linear model has several limitations include unprecise variance of correlation and potential biasness of coefficient. However, it is still useful in interpreting more than 50% of the variance of RDTP and shows the importance correlation of WTP with RDTP.

Business Implication

From the three models above, we can conclude that：

1. repurchase time do not affect per order transaction, therefore increasing repurchase time will increase revenue, given that number of consumers is stable.
2. consumers' satisfaction is negatively correlated to RDTP, higher RDTP can cause less repurchasing.
3. Waiting time proportion can affect RDTP in a significant level compared with other factors we proposed in the previous hypotheses.

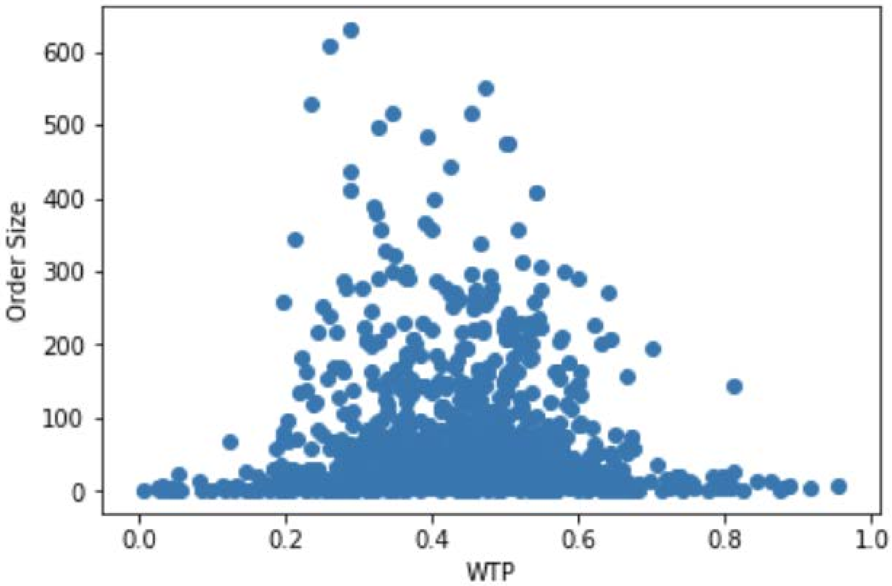
Therefore, to increase revenue by increasing repurchase times, we can implicate some strategies to impact RDTP by adjusting WTP.

Further Investigation for Strategy

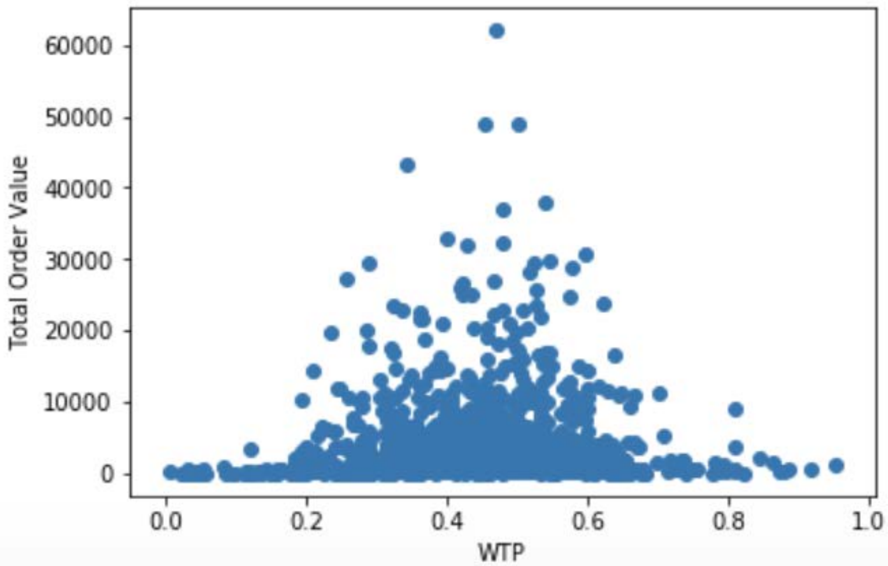
To find out what kind of strategies to implement, we study some characteristic of restaurants with high WTP. We find that restaurant with high WTP provide high order volume.

From figure 7 and figure 8, we can tell that restaurants with high WTP are responsible for certain amount of order size and revenue which is not low. Top 25% restaurants with highest WTP deliver 20% of orders (see appendix 6). Therefore, restaurants with high WTP play an important role on generating revenue and it is important to make strategies to improve their order performance.

Order Size Distributed along WTP [Figure 7]



Order Price Distributed along WTP [Figure 8]



Strategy to Reduce WTP

We conclude that the key to optimize on-time performance is to reduce WTP for restaurants with high WTP and also reduce the variance of WTP to make sure there is no significant order delay.

We discussed several options:

1. Improve recommendation system to let restaurants with long queues rank after restaurants with fewer order that normal condition.

2. Introduce score system which can reward the restaurants with good on-time performance and punish restaurants with bad on-time performance.

3. Reward customers who can order during low time. The three options focus on three different stakeholders in this system correspondingly.

After our careful analysis, we chose the first option and give up other two options. The reasons are as follows:

1. The ultimate goal is to improve revenue performance of the platform, if we give out subsidy or promotion to encourage customers order during low time, we will need to bear the financial burden. In fact, the subsidy we give out obviously cannot be offset by the gain of return purchase. It also decreases the customers' motivation to open the application for bulk/impulse purchase (e.g.: buy milk tea or fruits when ordering meals) during peak time.

2. The scoring system decreases the motivation of restaurants which have longer preparation time. In some cases, a restaurant which have long preparation time may be caused by an instantaneous peak demand or large order, which is more or less out of restaurants' control. It is unfair to punish restaurants in this situation. Restaurants will also notice the change, and some restaurants may leave the platform because of low score. It will decrease the volume of transactions since customers have fewer choices. What's more, rewarding restaurants with low WTP will cause more subsidies, which also betrays our ultimate goal.

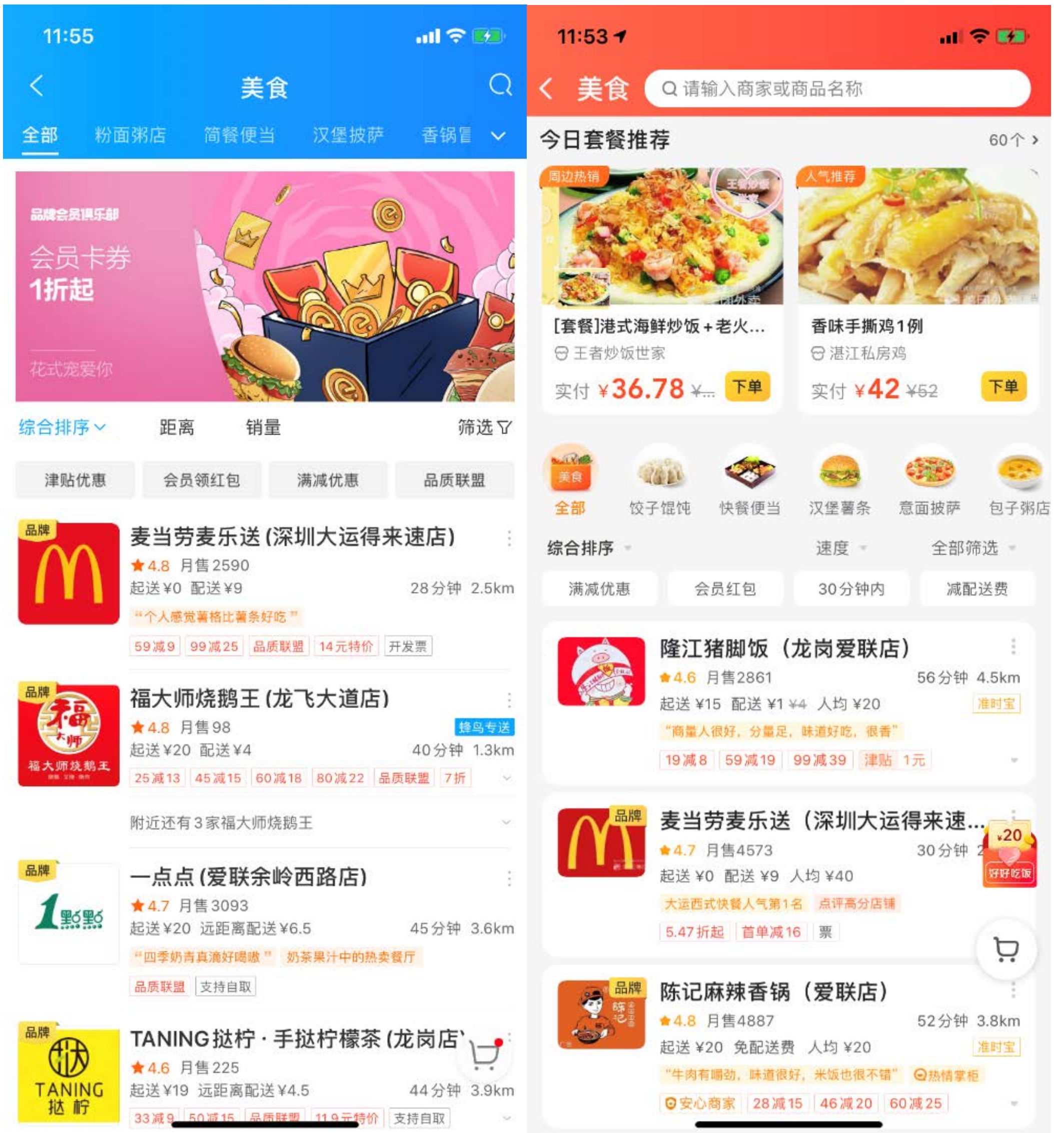
3. On the other hand, improved recommendation system can have the following advantages:

a. It is difficult for both restaurants and customers to figure out this change, and the churn rate will be significantly lower than scoring.

b. It does not cost much for the platform because it only requires minor adjustments without additional subsidies.

c. It can influence customers' choice in an indirect way, and also benefit restaurants because restaurants can predict the demand more accurately with more stable service quality and lower wastage of prepared food.

We can see from the following picture that both Eleme and Meituan's first page when ordering foods are comprehensive ranking （综合排序）. Comprehensive ranking relies on the algorithm of the platform to achieve more transactions. In some cases, the top restaurants shown on this page are paid listing. This page can greatly influence customers' choice because not all the customers are willing to search for foods they want.



(Left: Eleme. Right: Meituan Waimai)

By considering WTP in the recommendation system, restaurants which have instantaneously long waiting time or a large amount of orders will not be recommended to the first page, and restaurants which receive fewer order than normal days will be prioritized to shown on the first page. Restaurants which pay for the listing will stay on the same position regardless of its WTP. If customers filter based on other condition or use search function, the result remains the same. By implementing this change, restaurants who have more orders can have a higher chance to shorten WTP because there will be fewer incoming orders, and therefore the chance of delayed delivery is reduced and service quality is guaranteed. Restaurants with fewer orders than normal times will be encouraged to stay in the platform because of more incoming order. Since a restaurant may have its peak and low days and it's random, this change guarantees the fairness among restaurants. If a restaurant has high WTP over a long time, customers' comments will become more negative since more orders are delayed. It will also encourage restaurant to improve the WTP performance by adding capacity.

Conclusion

In this paper, we aim to address one pain point for meal delivery platform, the on-time performance, which is a crucial point for customers based on industry reports. To address on the on-time performance improvement, we raised two research question.

First is to investigate if the on-time performance has a significant impact on the customer satisfaction. Second is to further investigate how factors, including riders' waiting time, level of experience and efficiency of a particular rider, peek time orders, weather condition and weekdays orders, will impact the on-time performance.

To solve the research problem, three groups of corresponding hypotheses are raised. Then we used methods including correlation coefficient test, F-test and regression model to test the different hypothesis. Furthermore, to improve interpretation of the linear regression model, we test multicollinearity, heteroscedasticity and endogeneity problems and we find there is no severe multicollinearity, while other problems exist for the models and after trials, there is further need to address such problem in the future paper. However, it turns out to still be meaningful to interpret more than 50% of the variance of RDTP and shows the importance correlation of WTP with RDTP.

Therefore, one key finding of this paper is that, reducing WTP for restaurants with high WTP will optimize on-time performance and also reduce the variance of WTP to make sure there is no significant order delay. To utilize this finding and to provide a practical business strategy, we evaluated and compared three different solutions from meal delivery platform perspective. We concluded and suggested for the platform that improving recommendation system to let restaurants with long queues rank after restaurants with fewer order that normal condition is most applicable compare with other alternative solution. Lastly, the future direction for this paper is mainly about the improvements on models to better solve heteroscedasticity and endogeneity problems.

Reference

Yang, X. (2015). Chinese Internet Meal Delivery Industry analysis 2015. Retrieved 27 May 2020, from https://www.analysys.cn/article/analysis/detail/11671

Appendix

Appendix 1:

Result of correlation coefficient between repurchase time and per order transaction (the first number is the correlation coefficient at -0.068)

```
In [64]: corre = df_btest[['price','id']].corr().loc['price','id']
         print(corre)
         studentt2 = corre*(28495)**0.5/(1-corre**2)**0.5
         p = 1 - stats.t.cdf(abs(studentt2),df=28495)
         print(studentt2,p)
executed in 37ms, finished 09:27:20 2020-05-24

-0.06803843792369085
-11.511879184211695 0.0
```

Appendix 2:

RDTP and repurchase correlation coefficient (first number); RDTP and repurchase t-test result (second number)

```
In [208]: studentt = reg4.corr(method='pearson').loc['RDTP','id']*(reg4.shape[0]-2)**0.5/(1-float(reg4.corr(method='pe
          p = 1 - stats.t.cdf(abs(studentt),df=48)
          print(studentt,p)
          ##http://www.opentextbooks.org.hk/ditatopic/9498 How to calculate it
executed in 11ms, finished 00:06:50 2020-05-10
-1.0729970703744698 0.14431869165723643
```

Appendix 3:

correlation coefficient of working days and efficiency

	working_days	efficiency
working_days	1.000000	0.164962
efficiency	0.164962	1.000000

Appendix 4:

(p-value of the white test is the second number: 0.0)

```
(61754.424370466506, 0.0, 6368.493167909764, 0.0)
```

Appendix 5:

correlation coefficient of WTP and working day

```
-3.1940948497302107 0.0007017311302038687
-0.01202483731283216
executed in 12ms, finished 09:48:32 2020-05-24
print(studentt,p)
print(corre)
p = 1 - stats.t.cdf(abs(studentt),df=70542)
studentt = corre*(70542)**0.5/(1-corre**2)**0.5
In [84]: corre = df_m3[['RPP','working_days']].corr().loc['RPP','working_days']
```

correlation coefficient of WTP and order price

```
In [195]:  corre = df_m3[['price','PTP']].corr().loc['price','PTP']
           studentt2 = corre*(70545)**0.5/(1-corre**2)**0.5
           p = 1 - stats.t.cdf(abs(studentt2),df=70545)
           print(corre)
           print(studentt2,p)
```

executed in 16ms, finished 21:32:42 2020-05-25

```
0.12872908012643527
34.47770252270843 0.0
```

Appendix 6:

Top 25% restaurants with highest WTP deliver 20% of orders

```
In [200]:  df_sup.loc[(df_sup.PTP>0.518411),'id'].sum() / df_sup.loc[:,'id'].sum()
```

executed in 20ms, finished 21:59:02 2020-05-25

```
Out[200]: 0.20179225256799305
```