

For office use only

Team Control Number

For office use only

T1 \_\_\_\_\_

**2001560**

F1 \_\_\_\_\_

T2 \_\_\_\_\_

F2 \_\_\_\_\_

T3 \_\_\_\_\_

Problem Chosen

F3 \_\_\_\_\_

T4 \_\_\_\_\_

**C**

F4 \_\_\_\_\_

---

**2020**

**MCM/ICM**

**Summary Sheet**

## A Wealth of Data

### Summary

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Background . . . . .	3
1.2	Problem Restatement . . . . .	3
1.3	Data Source . . . . .	4
1.4	Data Pre-processing . . . . .	4
<b>2</b>	<b>Assumptions</b>	<b>4</b>
<b>3</b>	<b>Nomenclature</b>	<b>5</b>
<b>4</b>	<b>Model Design</b>	<b>6</b>
<b>5</b>	<b>Part I:</b>	<b>6</b>
<b>6</b>	<b>Part II:</b>	<b>6</b>
<b>7</b>	<b>Sensitivity Analysis</b>	<b>6</b>
<b>8</b>	<b>Strengths and Weaknesses</b>	<b>6</b>
8.1	Strengths . . . . .	6
8.2	Weaknesses . . . . .	6
<b>9</b>	<b>Conclusions</b>	<b>7</b>
	<b>Appendices</b>	<b>8</b>
	<b>Appendix A First appendix</b>	<b>8</b>
	<b>Appendix B Second appendix</b>	<b>8</b>

# LETTER

**To:** Marketing Director of Sunshine Company

**From:** MCM Team #2001560

**Subject:** Data Analysis Results

**Date:** March 8, 2020

---

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

# 1 Introduction

## 1.1 Background

In recent years, quantities of customers prefer to shopping online for its less spacetime limitation and the convenient home delivery service. However, compared to the traditional physical stores, customers can only evaluate products by the provided profile and pictures instead of seeing or testing the real ones. The information gap here is one of the leading causes of dissatisfied purchases. To help customers know the product better, many online marketplace platforms, such as Amazon, launch a "review system". Customers can express their level of satisfaction and further opinions or information about purchases through rating and reviewing. That additional information can help not only other customers make purchasing decisions, but companies improve the pros and cons of product design.

However, we found that not all reviews are equally relevant. Some reviews are too general; some people's ratings do not match their reviews; there are even deliberately misleading reviews, such as malicious defamation from competitors or the raise by the bribed reviewers. Therefore, when using data to assist business decisions, we need to analyze data carefully and comprehensively to obtain more accurate results. More factors should be considered, such as the ratings, review contents and review time, rather than straightly calculate the average rating level.

## 1.2 Problem Restatement

Analyze the three product data sets to describe quantitative and/or qualitative patterns, relationships that help evaluate a product's star ratings, reviews and helpfulness ratings. Use data to demonstrate that they are valuable.

Solve the following issues through modelling:

- Determine the most informative metric based on ratings and reviews. This metric can track the product ratings of three products when they are on the market.
- Analyze the relationship between product ratings and time in three data sets.
- Look for critical factors that can affect the inflexion point of product ratings through time.
- Analyze whether there will be more a series of positive or negative reviews over a while and whether customer star ratings will be affected by recent reviews.

- Whether star rating and the keyword of review content match.

### 1.3 Data Source

Our model is informed by the customer-supplied ratings and reviews for microwave ovens, baby pacifiers, and hair dryers sold in the Amazon marketplace over more than ten years.

### 1.4 Data Pre-processing

We did the following to sanitize the data set:

- Remove factors that were not measured at all, such as marketplace and product category, for they can not present any information.
- Remove the redundant factors, such as review\_id and product\_id, for they can be completely replaced by customer\_id and product\_parent.
- Remove factors that could mislead the model, such as verify\_purchase == N.
- Remove some garbled character.

## 2 Assumptions

- Merchant's purpose: Guide users to buy products with quality reviews and recent reviews.
- The content of the review and the star rating should be the same. People believe in reviews when review content is inconsistent with star ratings.  
Reason: Based on popular psychology, real language is convincing
- Amazon Vine members' reviews are credible, excluding subjective factors that give high ratings for free products  
Reason: Big data select Amazon Vine members. Their evaluations are more objective and practical.
- Actual purchasers' reviews are credible.  
Reason: People who have already experienced the product know more about the actual performance of the product.
- Comments from actual non-purchasers are untrustworthy.

Reason: We listed the distribution of 1-5 stars between the non-purchasers and the purchasers. It shows that the non-purchasers give more 1 star than purchasers, which may mislead the reviews.

- fig1+2

- The impact of the same purchaser on reviews is not taken into consideration; that is, each evaluation behaviour of the purchaser is independent and not related to the previous reviews.

### 3 Nomenclature

Symbol	Definition
$sr_i$	star rating of a review $i$
$L_i$	review length of a review $i$
$vh_i$	helpful votes of a review $i$
$vt_i$	total votes of a review $i$
$vm_i$	whether a review $i$ is from an Amazon vine member: yes, $vm_i=1$ ; no, $vm_i=0$ ;
$p_i$	whether a review $i$ is from a verified purchaser: yes, $p_i=1$ ; no, $p_i=0$ ;
$Vr_i$	votes rating of a review $i$
$Q_i$	review quality of a review $i$
$Q$	average review quality, can represent the review qualities in a certain time
$R_i$	review rating of a review $i$
$R$	average review rating, can represent the product ratings in a certain time
$t$	time

Table 1: variables and functions

## 4 Model Design

## 5 Part I:

## 6 Part II:

## 7 Sensitivity Analysis

## 8 Strengths and Weaknesses

### 8.1 Strengths

- **Applies widely**

This system can be used for many types of airplanes, and it also solves the interference during the procedure of the boarding airplane, as described above we can get to the optimization boarding time. We also know that all the service is automate.

- **Improve the quality of the airport service**

Balancing the cost of the cost and the benefit, it will bring in more convenient for airport and passengers. It also saves many human resources for the airline.

- 

### 8.2 Weaknesses and Improvement

- The small dataset causes deviation of the model which build by either word frequency analysis or LDA.

- The LDA model has an excellent performance with document collection, while the majority of reviews are short in words.

- While major reviewers are not Amazon Vine Member, different people may have different criteria for evaluation, which may produce an objective assessment. We need to study every reviewers' rating habit to get a weight for every reviewer.

- Removing the non-purchasers' reviews helps analysis the problem; however, these misleading reviews may hurt the customers' purchasing decisions and further ratings.

## 9 Conclusions

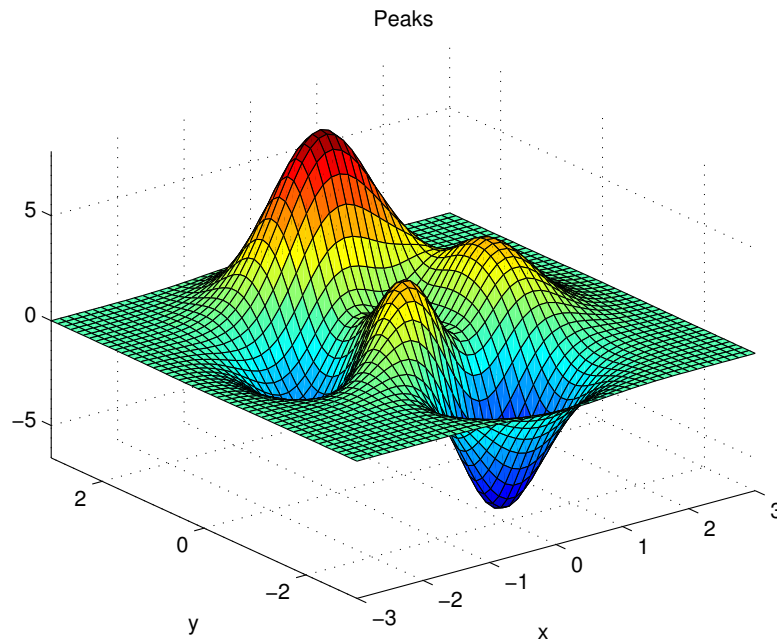


Figure 1: aa

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Donec odio elit, dictum in, hendrerit sit amet, egestas sed, leo. Praesent feugiat sapien aliquet odio. Integer vitae justo. Aliquam vestibulum fringilla lorem. Sed neque lectus, consectetur at, consectetur sed, eleifend ac, lectus. Nulla facilisi. Pellentesque eget lectus. Proin eu metus. Sed porttitor. In hac habitasse platea dictumst. Suspendisse eu lectus. Ut mi mi, lacinia sit amet, placerat et, mollis vitae, dui. Sed ante tellus, tristique ut, iaculis eu, malesuada ac, dui. Mauris nibh leo, facilisis non, adipiscing quis, ultrices a, dui. (1)

$$a^2 \quad (1)$$

$$p_j = \begin{cases} 0, & \text{if } j \text{ is odd} \\ r! (-1)^{j/2}, & \text{if } j \text{ is even} \end{cases}$$

## References

- [1] D. E. KNUTH The  $\text{\TeX}$ book the American Mathematical Society and Addison-Wesley Publishing Company , 1984-1986.
- [2] Lamport, Leslie,  $\text{\LaTeX}$ : " A Document Preparation System ", Addison-Wesley Publishing Company, 1986.



[3] <http://www.latexstudio.net/>

[4] <http://www.chinatex.org/>

# Appendices

## Appendix A First appendix

Aliquam lectus. Vivamus leo. Quisque ornare tellus ullamcorper nulla. Mauris porttitor pharetra tortor. Sed fringilla justo sed mauris. Mauris tellus. Sed non leo. Nullam elementum, magna in cursus sodales, augue est scelerisque sapien, venenatis congue nulla arcu et pede. Ut suscipit enim vel sapien. Donec congue. Maecenas urna mi, suscipit in, placerat ut, vestibulum ut, massa. Fusce ultrices nulla et nisl.

Here are simulation programmes we used in our model as follow.

### Input matlab source:

---

```
function [t,seat,aisle]=OI6Sim(n,target,seated)
pab=rand(1,n);
for i=1:n
    if pab(i)<0.4
        aisleTime(i)=0;
    else
        aisleTime(i)=trirnd(3.2,7.1,38.7);
    end
end
end
```

---

## Appendix B Second appendix

some more text **Input C++ source:**

---

```
//=====
// Name      : Sudoku.cpp
// Author     : wzlf11
// Version    : a.0
// Copyright  : Your copyright notice
// Description: Sudoku in C++.
//=====

#include <iostream>
#include <cstdlib>
#include <ctime>
```

```
using namespace std;

int table[9][9];

int main() {

    for(int i = 0; i < 9; i++){
        table[0][i] = i + 1;
    }

    srand((unsigned int)time(NULL));

    shuffle((int *)&table[0], 9);

    while(!put_line(1))
    {
        shuffle((int *)&table[0], 9);
    }

    for(int x = 0; x < 9; x++){
        for(int y = 0; y < 9; y++){
            cout << table[x][y] << " ";
        }

        cout << endl;
    }

    return 0;
}
```

---