

THE END OF 00 REPORT

SARAH SUN

ABSTRACT. Improve the online experience of potential customers by using the hypothetical organisation's hotel web log files to discover patterns of visits by different web users and to identify potential customers and their behavioural patterns.

CONTENTS

1. Introduction	2
2. Data cleaning	2
3. Data visualization	3
4. Conclusions	4

Date: 2021-04-23.

1991 *Mathematics Subject Classification.* TULIP.

Key words and phrases. L^AT_EX, Data visualisation, Exploratory data analysis.

TABLE 1. DATA ATTRIBUTE

Attribute Name	Description	Examples
date	Date of user access	1/11/2006
time	User access specific time	12:00:08 AM
s-ip	provides information on various protocols to the transport layer.	127.0.0.1
sc-status	Server successfully returns web page	200

```

date          0.000024
time          0.000024
s-sitename    0.000024
s-ip          0.000024
cs-method     0.000024
cs-uri-stem   0.000024
cs-uri-query  93.454194
s-port        0.000024
cs-username   100.000000
c-ip          0.000024
cs (User-Agent) 0.041818
cs (Referer)  15.519242
sc-status     0.008982
sc-substatus  0.008982
sc-win32-status 0.008982
dtype: float64

```

FIGURE 1. loading

1. INTRODUCTION

First of all, this hotel is a hypothetical organisation, like I just mentioned. It aims to train hoteliers, a hotel with creativity and the spirit of a TULIP lab. This hotel stores all and if the hotel website's web traffic and data about the various relevant web pages used by users.

2. DATA CLEANING

Extract the obtained data set: The obtained data set is in the form of a compressed package, which should be decompressed first. The names of the data sets in the log format of all the compressed packages shown in Figure 1 were obtained, and the statistics were carried out to find that there were 120 files in total.

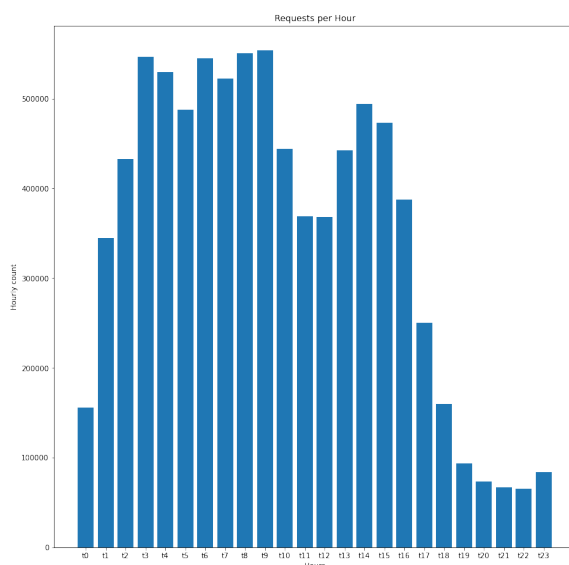


FIGURE 2. loading

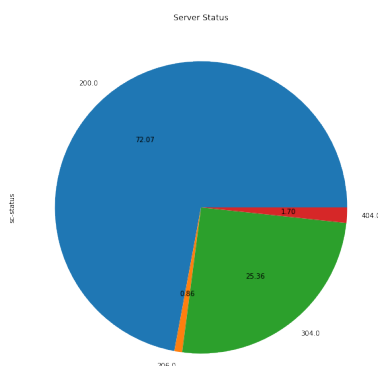
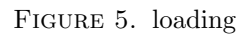


FIGURE 3. loading

3. DATA VISUALIZATION

First of all, I analyzed the click traffic. This is mainly for each hour to carry out statistics. Speaking each hour the same row counts as the flow plotted out. We can get some very interesting information. Due to the time difference, Melbourne time is adopted here, and we can see that there are more than 400,000 visitors in Melbourne time, and then there is a small peak from 2 to 9 in the morning, and then from 1 to 3 in the afternoon. There are also fewer visits between 7pm and 11pm

In the second step, I made statistics on the protocol status of the page visited by the user. Six million people were the most successful. In second place is rendering 304, which means the server has a cache for the web page. This proves that the user has visited the site for the second or more times.



and adjust the development strategy and planning of the organization in time.

(A. 1) SCHOOL OF BUSINESS AND LAW,, DEAKIN
UNIVERSITY,BURWOOD, VIC
Email address, A. 1: sunhan@deakin.edu.au

4. CONCLUSIONS

So just to recap a little bit, the reason I did all four of these analyses was to try to understand how often people were visiting over that time period, how often they were visiting, whether they were first-time visitors or repeat visitors. The main purpose of geographical analysis is to find out the geographical location of potential users, expand the market.