

# Hotel TULIP Web Server Data Analysis

**Assignment 2 - SIT742 Modern Data Science**

Alex Cummaudo <ca@deakin.edu.au>

Jake Renzella <renzella@deakin.edu.au>

Deakin Software and Technology Innovation Laboratory

School of Information Technology

Deakin University, Australia

April 28, 2017

## Executive Summary

This report summarises findings from a data exploration on the Hotel TULIP web server logs, recorded between the periods of August 2014 and August 2015. Each log contains one *request*, or *hit*, that lists fourteen attributes as described in the attached Data Dictionary spreadsheet. Publicly known client IP addresses were extracted from the MaxMind GeoIP2<sup>1</sup> dataset to analyse the location of requests (narrowed down to city). Additionally, user agent strings were parsed to analyse device and browser statistics using the Python user-agents library<sup>2</sup>, thereby extrapolating demographics, usage trends, platform information, server performance, and security statistics from the raw logs provided in the dataset. Further details on the extraction of the data is provided in the source code attached in Appendix B, and an interactive version of this file is published on Databricks.

---

<sup>1</sup>See <http://dev.maxmind.com/geoip/geoip2/>.

<sup>2</sup>See <https://pypi.python.org/pypi/user-agents>.

# Contents

<b>1</b>	<b>Key Findings</b>	<b>7</b>
<b>2</b>	<b>Demographics</b>	<b>8</b>
2.1	Countries . . . . .	8
2.2	Cities . . . . .	8
2.3	Request Sources . . . . .	8
<b>3</b>	<b>Usage Trends</b>	<b>10</b>
3.1	Overall usage . . . . .	10
3.2	Hourly Hits . . . . .	10
3.3	Daily Hits . . . . .	10
<b>4</b>	<b>Platforms &amp; Operating Systems</b>	<b>17</b>
4.1	Smartphone and Tablet Adoption . . . . .	17
4.2	PC Adoption . . . . .	24
4.3	Device Categories . . . . .	25
4.4	Browser Usage . . . . .	27
<b>5</b>	<b>Server Performance</b>	<b>30</b>
5.1	Response Times . . . . .	30
5.2	Security . . . . .	30
5.3	Server Errors . . . . .	35
<b>6</b>	<b>Technical</b>	<b>38</b>
6.1	SEO Traffic . . . . .	38
6.2	Resources Accessed . . . . .	38
6.3	Methods Used . . . . .	38
<b>7</b>	<b>Additional Points of Interest</b>	<b>41</b>
7.1	Username . . . . .	41
7.2	Frequent Clients . . . . .	41

7.3	Query Strings . . . . .	41
7.4	Server IP Usage . . . . .	41
<b>A</b>	<b>Additional Figures</b>	<b>42</b>
<b>B</b>	<b>Extrapolation Results</b>	<b>77</b>

## List of Figures

1	Proportion of requests made from top ten countries . . . . .	11
2	Requests made by top four and ten countries over 12 months . . . . .	12
3	Locations of cities where requests were made within the top four countries . . . . .	13
4	Trends of hits over 12 months . . . . .	14
5	Trends of hits per hour . . . . .	15
6	Trends of hits each month . . . . .	16
7	Proportion of operating system usage by top ten countries . . . . .	17
8	Operating system proportions of requests made both internally and externally . . .	18
9	Trend of internal requests made by iOS versus Android . . . . .	19
10	iPad versus iPhone requests . . . . .	21
11	Device brand distribution between internal and external requests . . . . .	22
12	Proportion of device brands used in requests by top ten countries . . . . .	23
13	Trend of usage of Windows, Mac OS X and Linux of requests made internally . . .	24
14	Trend of requests made by PCs, tablets, smartphones and bots . . . . .	25
15	Proportion of requests made by PCs, tablets, smartphones and bots by top ten countries . . . . .	26
16	Browser usage distribution between internal and external requests . . . . .	28
17	Proportion of browsers used in requests by top ten countries . . . . .	29
18	Average response time trends over previous twelve months . . . . .	31
19	Average response times by country . . . . .	32
20	Number of times <i>Forbidden</i> request attempts were made and when . . . . .	33
21	Number of times <i>Forbidden</i> request attempts were made by country and proportions of humans versus bots making those attempts . . . . .	34
22	Number of server errors made in the last twelve months . . . . .	36
23	Number of server errors caused by bots versus humans per country . . . . .	37
24	Proportions of bots that crawled the server . . . . .	39
25	Number of requests made by varying bots per country . . . . .	40
26	Trend of requests made by varying mobile operating systems . . . . .	42
27	Trend of requests made by varying PC operating systems . . . . .	43

28	Trend of requests made by varying mobile operating systems (excluding iOS and Android) . . . . .	44
29	Most frequently returning visitors and respective cities . . . . .	45
30	Proportion of internal versus external requests made . . . . .	46
31	Trend of requests made by varying PC operating systems (excluding Windows and Mac OS X) . . . . .	47
32	Requests trends of device brands over time . . . . .	48
33	Trends of device brand usage over time (without Apple) . . . . .	49
34	Proportion of Mac OS X versions used to make requests . . . . .	50
35	Proportion of Mac OS X versions distributed by top ten countries . . . . .	51
36	Proportion of requests made by varying archaic Windows versions . . . . .	52
37	Proportion of requests made by varying archaic Windows versions by country . . . . .	53
38	Proportions of resources accessed by internal and external requests . . . . .	54
39	Proportions of human-accessed resources by internal and external requests . . . . .	55
40	Proportions of resources requested by bots only . . . . .	56
41	Proportions of query strings used in all requests with query strings . . . . .	57
42	Proportions of HTTP methods used in each request . . . . .	58
43	Proportions of HTTP non-GET methods used in each request . . . . .	59
44	Aggregated hourly trends of HTTP requests with POST method . . . . .	60
45	Proportions of all HTTP response codes . . . . .	61
46	Proportions of all non-200 HTTP response codes . . . . .	62
47	Number of aggregated server errors made per day in each month . . . . .	63
48	Number of aggregated <i>Forbidden</i> requests per day in each month . . . . .	64
49	Resources with fastest average response times . . . . .	65
50	Resources with slowest average response times . . . . .	66
51	Top frequently accessed resources and their respective response times . . . . .	67
52	Average response time by operating system . . . . .	68
53	Number of times varying ports accessed . . . . .	69
54	Most frequently accessed resources over HTTPS . . . . .	70
55	Number of aggregated HTTPS requests per hour . . . . .	71

56	Number of aggregated HTTPS requests per hour . . . . .	72
57	Proportion of non-empty usernames used . . . . .	73
58	Proportion of varying Win32 status codes . . . . .	74
59	Proportion of varying IIS substatus codes against HTTP status code . . . . .	75
60	Proportion of server IP usage . . . . .	76

# 1 Key Findings

A list of key findings in the analysis are as thus:

- A majority of requests are from Hong Kong, specifically the Central District (Section 2.1).
- The number of requests made to the web server are increasing over time (Section 3).
- Response times to the server are getting slower (Section 5.1).
- There are a number of increasing server errors (Section 5.3).
- The web server is most active at 9am UTC, with most requests made in March 2015 (Section 3).
- iOS is being used more than Android (Section 4.1).
- The server is generally secure with relatively few genuine repetitive unauthorised attempts (Section 5.2).
- iPhone is being used more than iPad to request the server worldwide, but internally, guests use iPad more than iPhone (Section 4.3)
- Windows was mostly used for requests made on PC devices (Section 4.2).
- Apple devices are primarily used around the world for smartphone and tablet requests (Section 4.1).
- Google is crawling at 16%, over Bing and Facebook at 10% and 4% (Section 6.1).



## 2 Demographics

### 2.1 Countries

Figure 1 highlights the proportions of countries that have made at least 10,000 requests to the hotel's server. The maximum number of requests are from Hong Kong (56% or 39.97m), 8% (5.52m) from the United States, 5% (3.23m) from Australia and 4% (2.70m) from the United Kingdom. 16% are from other countries.

From Figure 2, we see a growing trend of Hong Kong visitors to the website, though there has been some decline since March 2015. Visitors from western countries (here, Australia, the U.K. and U.S.) grew from August 2014 and peaked around January to March 2015 (315,940 from Australia; 272,741 from the U.K., and 529,990 from the U.S., respectively). Japanese and Taiwanese visitors have largely been steady, fluctuating at around 100,000 requests, whereas Chinese and Korean visitors have spiked between June and August 2016 to 293,089 and 244,621 hits, respectively.

### 2.2 Cities

We analysed from these top four countries where the majority of requests were coming from within each city (Figure 3). We see that:

- The majority of Hong Kong visitors are from within the Central District, at 77% (27.36m),
- 28% (880k), 8% (240k), 5% (162k) of American visitors are from Mountain View, New York and Redmond, respectively,
- The majority of U.K. visitors are from London, at 40% (346k), and
- 23% (240k), 15% (158k), and 12% (129k) of Australian visitors are from Sydney, Melbourne and Perth, respectively.

### 2.3 Request Sources

An *internal request* is a request that was made from within the Hotel TULIP network. These requests may be made by guests, those connected whilst in the lobby, staff, and so on. We compare

the amount of requests made internally versus externally in Figure 30; 1,774,018 internal requests were made within the period, when compared to 71 million external requests.

## 3 Usage Trends

### 3.1 Overall usage

Figure 4 shows a decline in the number of internal hits in the website since August 2014, with a peak in September 2014 at 176k hits down to a trough of 103k in February 2015. There was improvement since, inclining to 121k, but this is still a 31% decrease in hits in the timespan listed. Investment into improving the internal networks may not be necessary should this declining trend continues.

The positive side is an overall increase in visits to the website worldwide, increasing from 4.6 million in August 2014 to 5.8 million a year later. We see the busiest time in the calendar year is around March period, with almost 7 million requests made.

### 3.2 Hourly Hits

From Figure 5, we see the website generally was most busy between 3am and 9am (UTC), though there is a second local maximum for non-internal requests at around 2pm UTC. This is not consistent with the internal requests made, as it is generally quiet between 12pm and 9pm UTC. Consideration into the fact that the time is measured in UTC should be accounted for.

Requests made over HTTPS are shown within Figure 55, and tends to be consistent with the hit pattern observed in the non-HTTPS requests.

### 3.3 Daily Hits

The website is shown to generally increase in the number of hits as the month progresses (Figure 6), although this trend is not observed for internal requests only. Internal requests fluctuate, and generally are shown to trough around every seven days in the month.

figs/4\_1\_top\_countries.pdf

Figure 1: Proportion of requests made from top ten countries where 10,000 requests were made

figs/4\_3\_top\_cities\_trends.pdf

figs/4\_3\_top\_cities\_trends\_without\_hk.pdf

figs/4\_2\_top\_cities.pdf

Figure 3: Locations of cities where requests were made within the top four countries

figs/6\_4\_hits\_trends\_by\_all.pdf

figs/6\_4\_hits\_trends\_by\_guests.pdf

figs/6\_1\_hits\_trends\_by\_hour.pdf



figs/6\_2\_hits\_trends\_by\_month.pdf

figs/6\_3\_hits\_trends\_by\_month\_guests.pdf

## 4 Platforms & Operating Systems

figs/5\_3\_1\_os\_type\_by\_country.pdf

Figure 7: Proportion of operating system usage by top ten countries

### 4.1 Smartphone and Tablet Adoption

Most internal requests within the hotel were made on iOS devices, with relatively few on made on Android. We can compare the peak of iOS's maximum at 2.9k with Android's 233 made at January

figs/5\_3\_1\_os\_type.pdf

Figure 8: Operating system proportions of requests made both internally and externally

figs/5\_3\_3\_ios\_vs\_android\_used\_by\_guests.pdf

Figure 9: Trend of internal requests made by iOS versus Android

2015 and July 2015. This is outlined in Figure 9. We therefore can assume that there is a larger proportion of iOS users than Android users within the hotel.

Similarly, there is a marginal lead by iOS over Android for all requests made within Figure 26 by a consistent approximate of 700k requests. A comparison of requests of non-iOS and Android mobile OSes are given in Figure 28.

From Figure 7, we see that Android is most used in Hong Kong at about 13%. iOS users are more prevalent in Australia and the U.K. at about 25%.

In terms of total proportion, Figure 8 shows us that most mobile-requests came externally from iOS (14.2m or 20%), whereas about half of this (6.8m or 10%) are from Android.

Additionally, we can compare the brand of smartphones and tablets used in the internal and external requests. From Figure 11, we see that Apple devices are primarily used for both (67% externally, 91% internally). This is consistent with usage from the top ten countries (Figure 12), the highest Apple users being in Australia and the U.K., closely followed by Japan. Samsung devices are the next major brand to follow consistently in both areas.

iPad and iPhone usage has fluctuated—Figure 10 shows that guests prefer to use iPads over their iPhones when within the hotel. However, iPhone clearly outperforms the iPad (since November 2014) for all requests made; more than double the hits are from iPhone than iPad.

figs/5\_2\_3\_ipad\_vs\_iphone\_internal.pdf

figs/5\_2\_1\_device\_dist.pdf

Figure 11: Device brand distribution between internal and external requests

figs/5\_2\_1\_device\_dist\_by\_country.pdf

Figure 12: Proportion of device brands used in requests by top ten countries



## 4.2 PC Adoption

Guests largely within the hotel's network are accustomed to using Windows on their desktops, though there has been a steady decline from 137k to 119k requests over the twelve month period. See Figure 13.

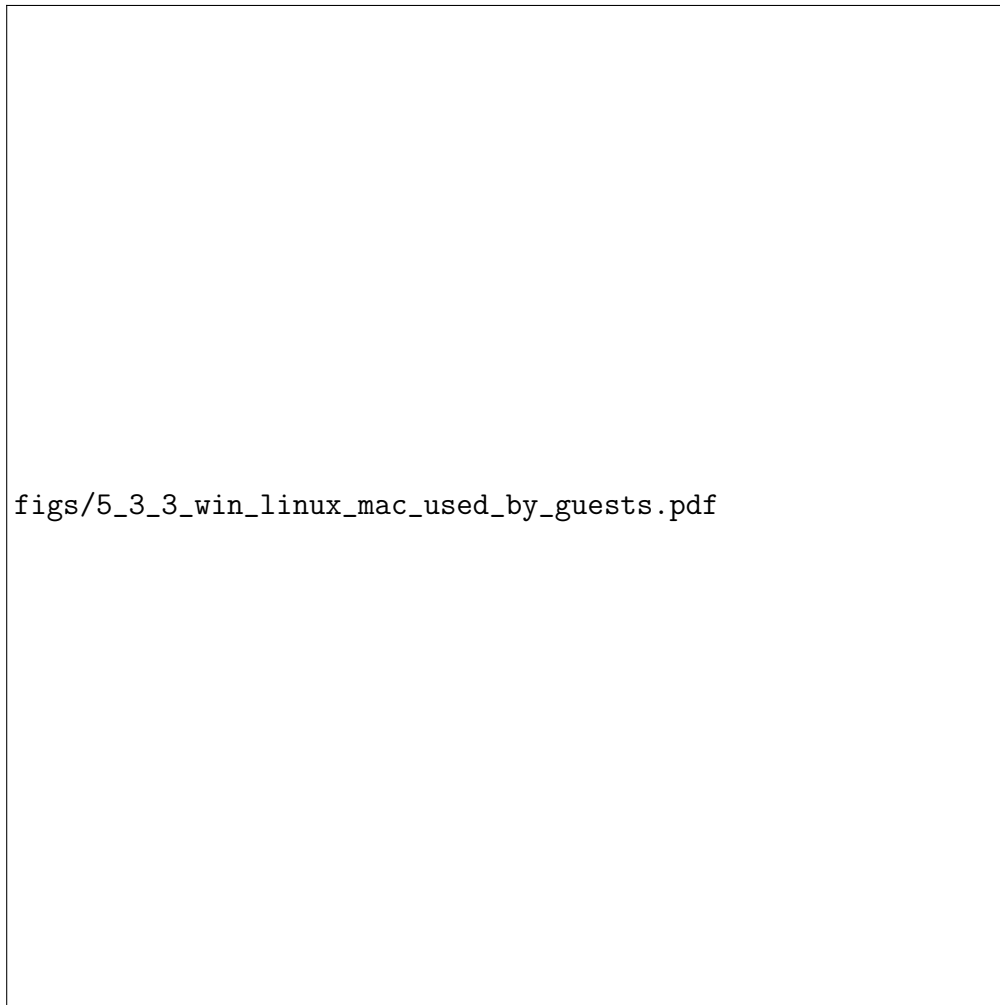


Figure 13: Trend of usage of Windows, Mac OS X and Linux of requests made internally

Requests made to the network in general were largely from Windows-based machines (Figure 27), the highest proportion of Windows users being in the Republic of Korea at about 75% (Figure 7). The highest proportion of Mac users were from Australia, the U.K., then the U.S.

There is still adoption of archaic version of Windows<sup>3</sup>, particularly in Hong Kong, China and

---

<sup>3</sup>Modern versions of Windows beyond Windows 2000 cannot be determined for technical reasons. Refer to the notebook for more.

(to some extent) the U.S. (Figure 37). These versions are outlined within Figure 36. Similar results for Mac OS X versions are shown in Figures 35 and 34, respectively.

### 4.3 Device Categories



Figure 14: Trend of requests made by PCs, tablets, smartphones and bots

As outlined in Figure 14, we see a rapid increase of smartphone usage accessing the website since September 2014, where smartphones outpaced both PCs and Tablets after April 2015. Tablet

figs/5\_2\_4\_smartphone\_tablet\_pc\_bots\_comparison\_by\_country.pdf

Figure 15: Proportion of requests made by PCs, tablets, smartphones and bots by top ten countries

and PC usage maxed in January 15 and have been on a steady decline since.

We can breakdown usage in the top ten countries outlined in Figure 15: by proportion of requests, PCs are mostly used within Australia at about 45% of all the country's requests. This is followed closely by the U.S. and U.K., which have comparable proportions. Hong Kong and the Republic of Korea have similar proportion of smartphone users, at about 40%. The U.K. has the highest proportions of tablet users at 43%, followed next by Australia at 38%. Bots are only used from China and the U.S. in comparable percentages of approximately 15% of all requests from their respective countries.

## **4.4 Browser Usage**

From Figure 16, 18% of users (327.5k) used Google Chrome within the Hotel's network, compared to 35% of users (25.25m) using Chrome externally. This was followed closely by Firefox (3%) within the internal network, but externally the second-most popular browser was Mobile Safari (17%). While 11% of requests came from Internet Explorer, it is most popular in Korea (Figure 17), whereas Chrome is the browser of choice elsewhere (except Japan, which has an even distribution between Chrome and Internet Explorer).

figs/5\_1\_browser\_dist.pdf

Figure 16: Browser usage distribution between internal and external requests

figs/5\_1\_browser\_dist\_by\_country.pdf

Figure 17: Proportion of browsers used in requests by top ten countries

## 5 Server Performance

### 5.1 Response Times

Average response times (as measured in milliseconds) are generally increasing over time (Figure 18). This is inline with the increasing trend of requests shown in Figure 4. Improvement into server response times may be required if traffic is to increase, slowing processing speeds.

On a per-country basis, requests from the top ten fastest countries can usually respond in under 900ms, averaging at about 550ms. Hong Kong, Singapore and the Republic of Korea are usually the fastest, with average responses in under 400ms. The inverse of this shows Nigeria as the slowest, taking on average 4.5 seconds.

The statistics shown in Figure 19 show these fastest and slowest countries, granted that at least 10,000 requests have been made.

On average, Windows is the fastest to operating system to respond to, followed by Linux and Mac OS X. For mobile operating systems, iOS typically responds in about 700ms, while Android is slower at 880ms. This data is extrapolated from Figure 52.

### 5.2 Security

A number of requests returned a HTTP 403 (*Forbidden*) response code. Most attempts at accessing unauthorised files were made at around 6am UTC (Figure 20). Most of these requests, as shown in Figure 48, were made in August, October and September 2014, and none were reported since.

It is worth noting whether or not these forbidden requests were made by humans (attempted hacks) or bots (attempted crawling). While most forbidden requests were sourced from the U.S. and China, only 22 were made by humans in these countries. Only Hong Kong had human-only forbidden requests at 65 forbidden requests. A majority of the requests (203) were made by bots, versus 101 made by humans. This is outlined within Figure 21.

Figure 51 shows a good response time for the most frequently accessed resource (About Us page in under 400ms, on average), though the Home page performance could be improved at an average response time of 2.2 seconds. The following four most access resources average at about a second in their response times.

HTTP requests typically outperform HTTPS responses by a fraction of 20% or about 200ms,


figs/8\_2\_3\_response\_time\_trends.pdf

Figure 18: Average response time trends over previous twelve months



figs/8\_2\_1\_fastest\_response\_times.pdf

figs/8\_2\_2\_slowest\_response\_times.pdf



figs/8\_1\_2\_forbidden\_attempts\_all\_time.pdf

Figure 20: Number of times *Forbidden* request attempts were made and when

figs/8\_1\_2\_forbidden\_attempts\_by\_humans\_vs\_bots.pdf

Figure 21: Number of times *Forbidden* request attempts were made by country and proportions of humans versus bots making those attempts

as shown in Figure 56.

### 5.3 Server Errors

The number of errors<sup>4</sup> being made are increasing with a steady incline, from 500 in August 2014 up to 4,750 a year later. This is plotted in Figure 22. This 89% increase in errors may also be due to an increase in requests (Figure 4). It is notable that a majority of the requests 15k errors are caused by humans in Hong Kong, whereas 7k are by bots from the United States. Figure 23 notes this in further detail.

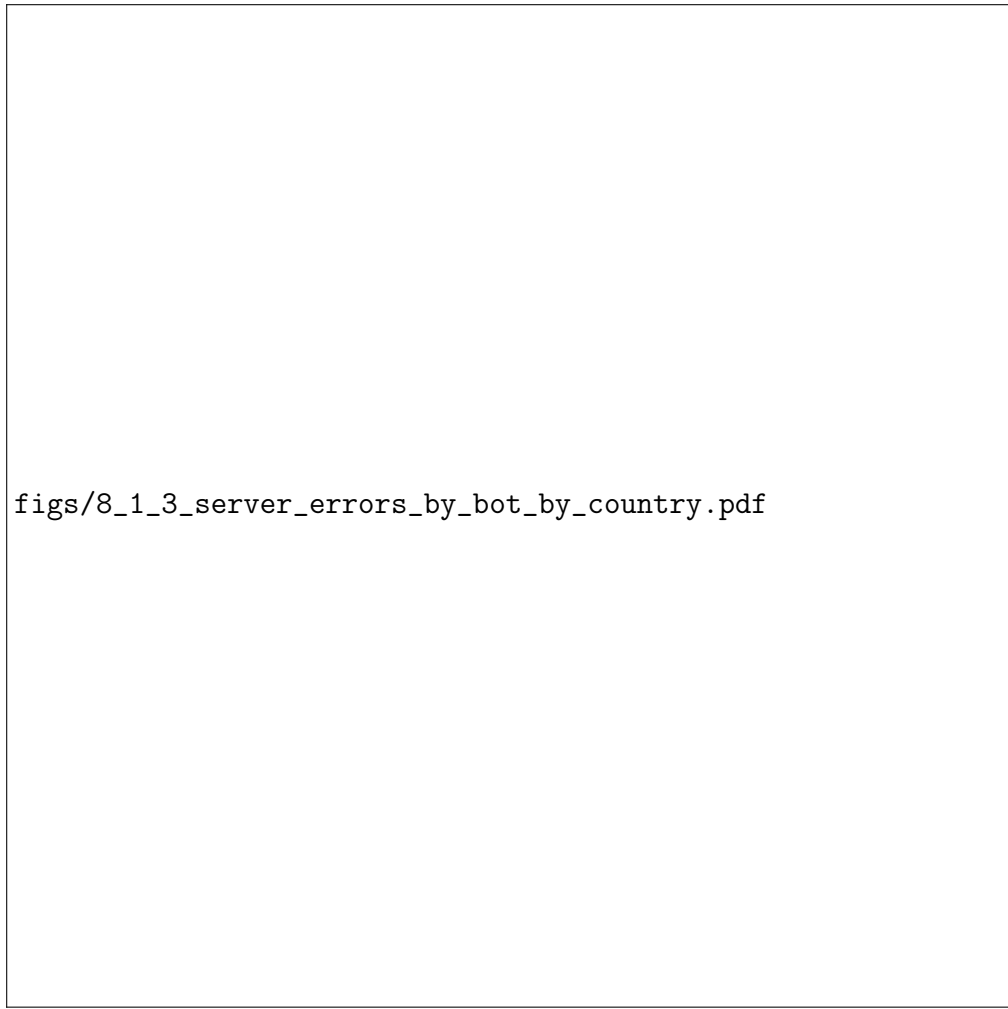
Refer to Figures 58 and 59 for more detailed error codes reported. Files not being found or network names being removed are the cause of most Win32 errors, at 53% and 44%, respectively. The largest IIS subcode reported for *Forbidden* response codes (403) was the directory listing could not be presented to the user (403.14), and with *Not Found* response codes (404), the largest was that a MIME type restriction occurred (404.3).

---

<sup>4</sup>Where a server error is defined as a HTTP 50x response code.

figs/8\_1\_3\_server\_errors\_all\_time.pdf

Figure 22: Number of server errors made in the last twelve months



figs/8\_1\_3\_server\_errors\_by\_bot\_by\_country.pdf

Figure 23: Number of server errors caused by bots versus humans per country

## 6 Technical

### 6.1 SEO Traffic

From Figure 24, we are able to determine that:

- 39% of traffic is crawled by *PingdomBot*<sup>5</sup>, a free website availability and performance bot,
- 16% of traffic is crawled by Google, and
- 10% of traffic is crawled by Bing

The majority and diversity of bots are from the United States, as shown in Figure 25. China exclusively made bot requests using the *Baiduspider*<sup>6</sup> bot, whereas Russia used *YandexBot*<sup>7</sup>. Elsewhere in the world, *PingdomBot* is used.

### 6.2 Resources Accessed

Figures 38, 39 and 40 show all resources accessed, those accessed by non-bots and those accessed by bots. 28% of external requests are for the ‘About Us’, compared to 18% of internal requests. Of the dining and offers pages, only 11% and 7% of requests are made to these resources. Internal requests primarily source to administration controls, such as the `sitecore/shell` directory.

Bots make a significant amount of crawling on the hotel’s ‘Wine and Dine’ page, at 5%, though a majority are at the index page at 64%. The bots only crawl the `robots.txt` file at 4% of the time.

The fastest responses made by the server on resources are the `autodiscover.xml` file, at just 53ms. Themes and fonts come next, ranging from 55ms to 80ms. Further detail is shown in Figure 49. The slowest response comes from the rooms PDF, which takes, on average 11 seconds to download. Refer to Figure 50.

### 6.3 Methods Used

Almost exclusively requests are GET requests (Figure 42), though when we do not consider this outlier, 68% of requests (141k) are POSTing to the server (Figure 43). We can see that most users

---

<sup>5</sup>See <https://www.pingdom.com/>

<sup>6</sup>See <https://chineseseoshifu.com/blog/what-is-baiduspider.html>

<sup>7</sup>See <https://yandex.com/bots>

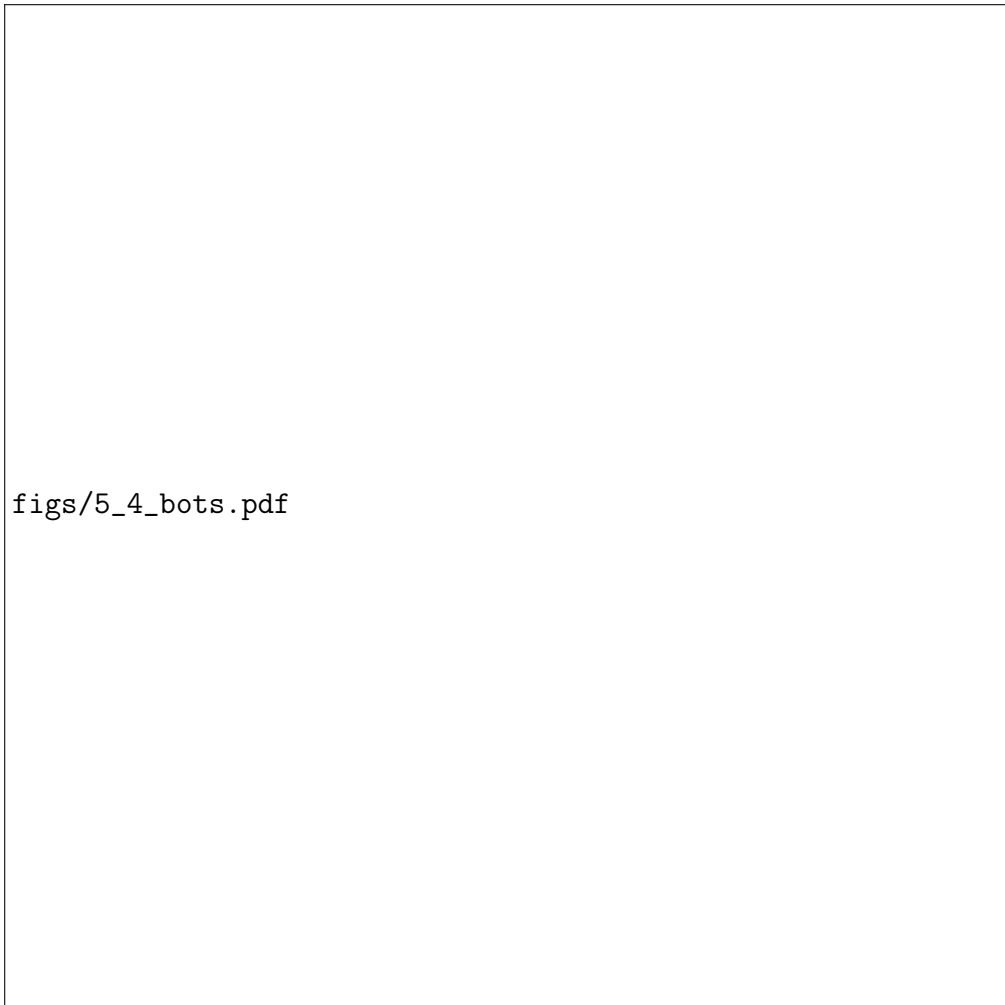


Figure 24: Proportions of bots that crawled the server



figs/5\_4\_bots\_by\_country.pdf

Figure 25: Number of requests made by varying bots per country

will tend to POST data at around 1pm to 3pm UTC, as shown in Figure 44.

## **7 Additional Points of Interest**

### **7.1 Usernames**

While most usernames are not provided in a majority of requests, we can deduce from Figure 57 that the Administrator accessed the website, when compared to other users, 76% of the time. This translates to 1,513 requests. User ‘Alex Islam’ came second at 11% of requests, or 216 requests, closely followed by ‘Human Capital’ at 10% of requests at 207 requests.

### **7.2 Frequent Clients**

The most frequent clients are distributed in Figure 29. We can visualise where the 10 most frequent visitors are from, based on their IPs. The top three are all from three clients in Hong Kong:

1. 3,474,747 hits are from an IP 14.136.194.139 in the Central District,
2. 102,536 hits are from an IP 123.203.152.174 in Kowloon, and
3. 89,532 hits are from an IP 203.90.7.79 in the Central District

There is one outlier: a frequent visitor from Philadelphia that returned 54,183 times. This was the 8th most frequent visitor.

### **7.3 Query Strings**

Query strings are shown in Figure 41. While usually blank, we see that 35% of a particular query string is given in over 1 million requests to the site.

### **7.4 Server IP Usage**

Almost all requests were made to the IP 10.130.0.12, as shown in Figure 60.

## A Additional Figures

figs/5\_3\_3\_mobile\_os.pdf

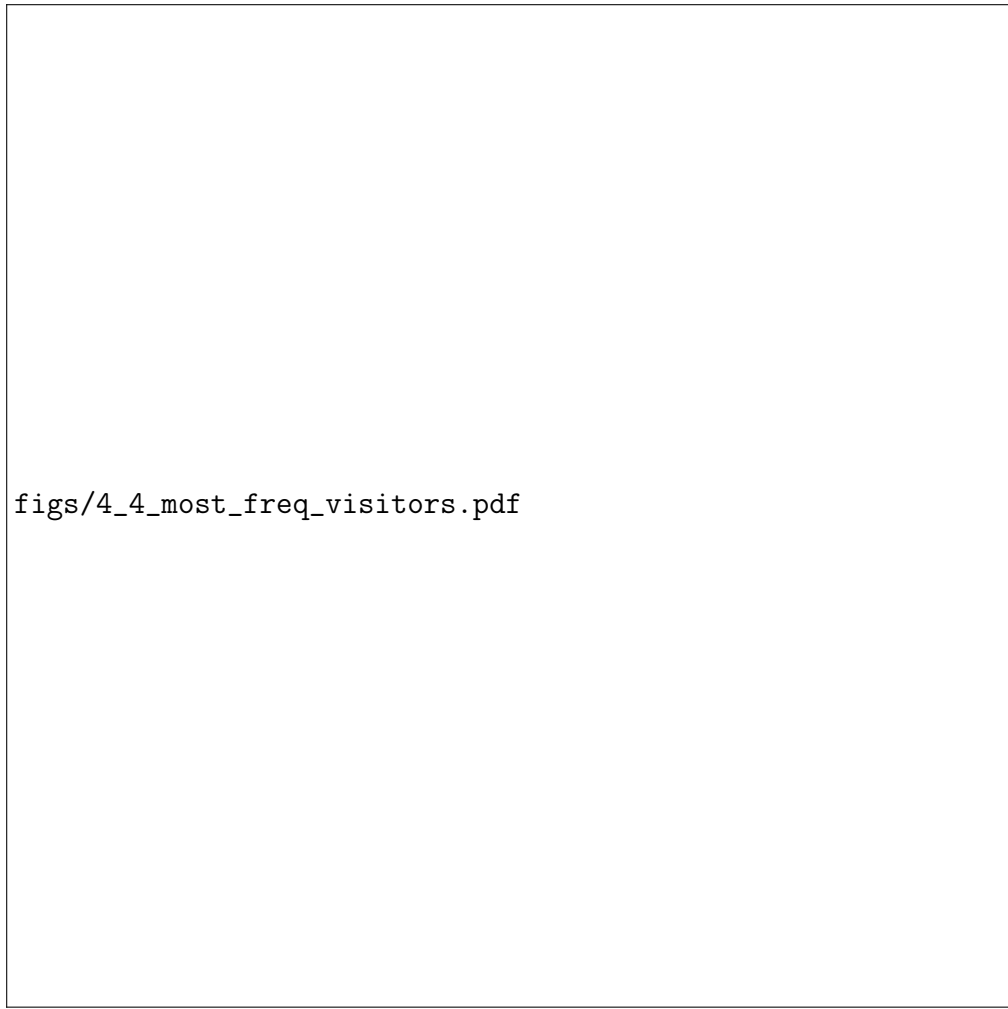
Figure 26: Trend of requests made by varying mobile operating systems

figs/5\_3\_3\_pc\_os.pdf

Figure 27: Trend of requests made by varying PC operating systems

figs/5\_3\_3\_mobile\_os\_no\_apple\_or\_android.pdf

Figure 28: Trend of requests made by varying mobile operating systems (excluding iOS and Android)



figs/4\_4\_most\_freq\_visitors.pdf

Figure 29: Most frequently returning visitors and respective cities

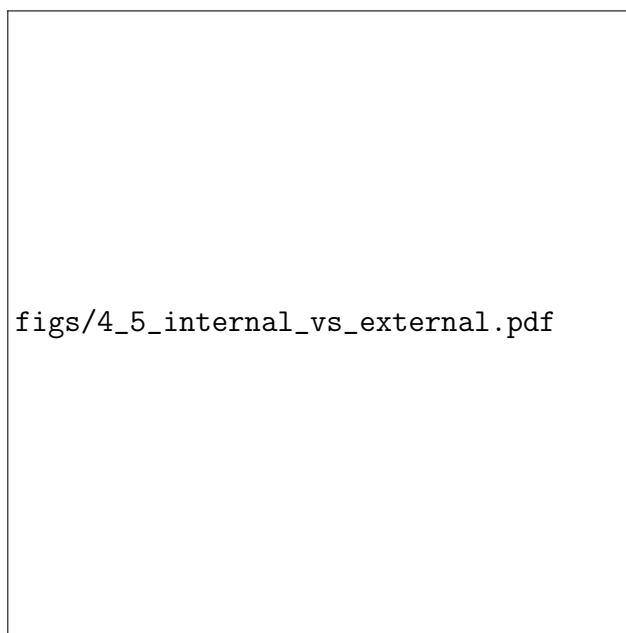


Figure 30: Proportion of internal versus external requests made

figs/5\_3\_3\_pc\_os\_no\_mac\_win.pdf

Figure 31: Trend of requests made by varying PC operating systems (excluding Windows and Mac OS X)

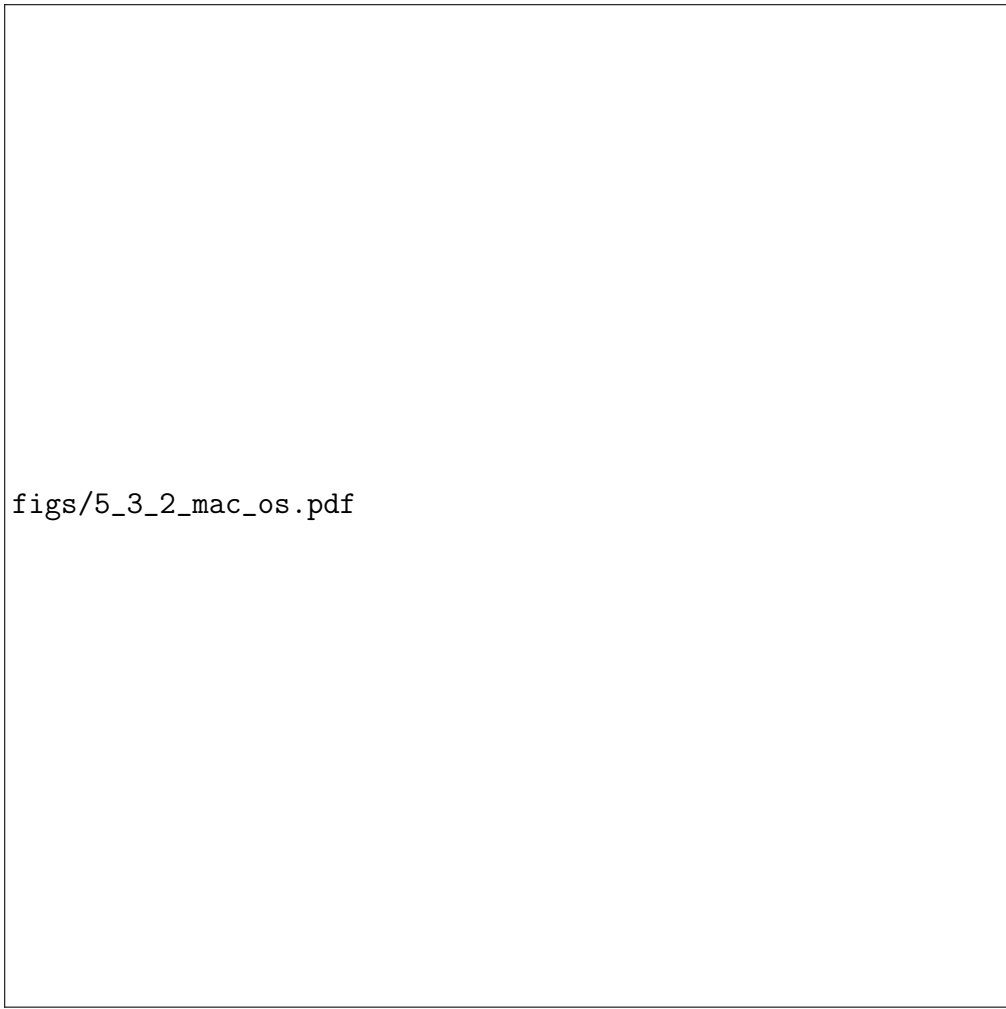


figs/5\_2\_2\_device\_adoption.pdf

Figure 32: Requests trends of device brands over time

figs/5\_2\_2\_device\_adoption\_no\_apple.pdf

Figure 33: Trends of device brand usage over time (without Apple)

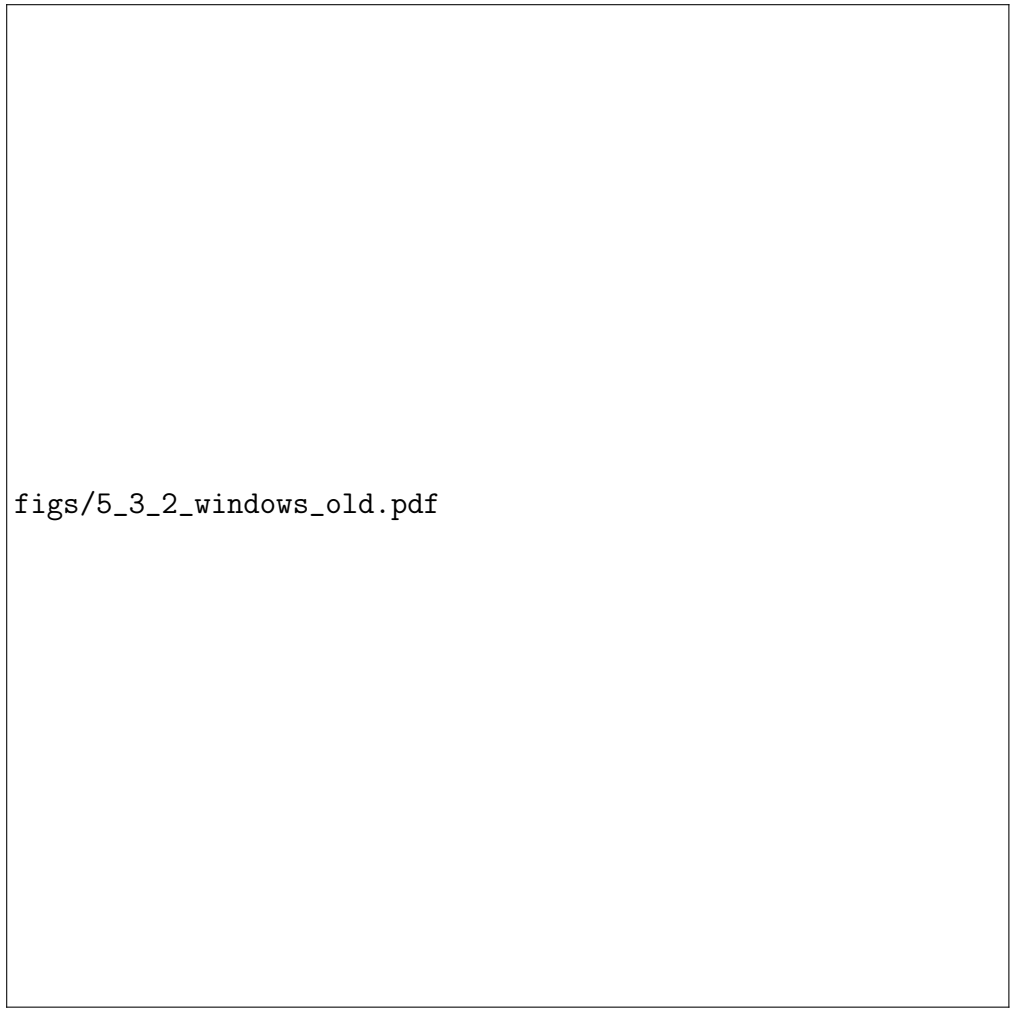


figs/5\_3\_2\_mac\_os.pdf

Figure 34: Proportion of Mac OS X versions used to make requests

figs/5\_3\_2\_mac\_os\_by\_country.pdf

Figure 35: Proportion of Mac OS X versions distributed by top ten countries



figs/5\_3\_2\_windows\_old.pdf

Figure 36: Proportion of requests made by varying archaic Windows versions

figs/5\_3\_2\_windows\_old\_by\_country.pdf

Figure 37: Proportion of requests made by varying archaic Windows versions by country


figs/7\_1\_1\_resources\_accessed.pdf

Figure 38: Proportions of resources accessed by internal and external requests

figs/7\_1\_2\_resources\_accessed\_humans\_only.pdf

Figure 39: Proportions of human-accessed resources by internal and external requests





figs/7\_1\_3\_resources\_accessed\_bots\_only.pdf

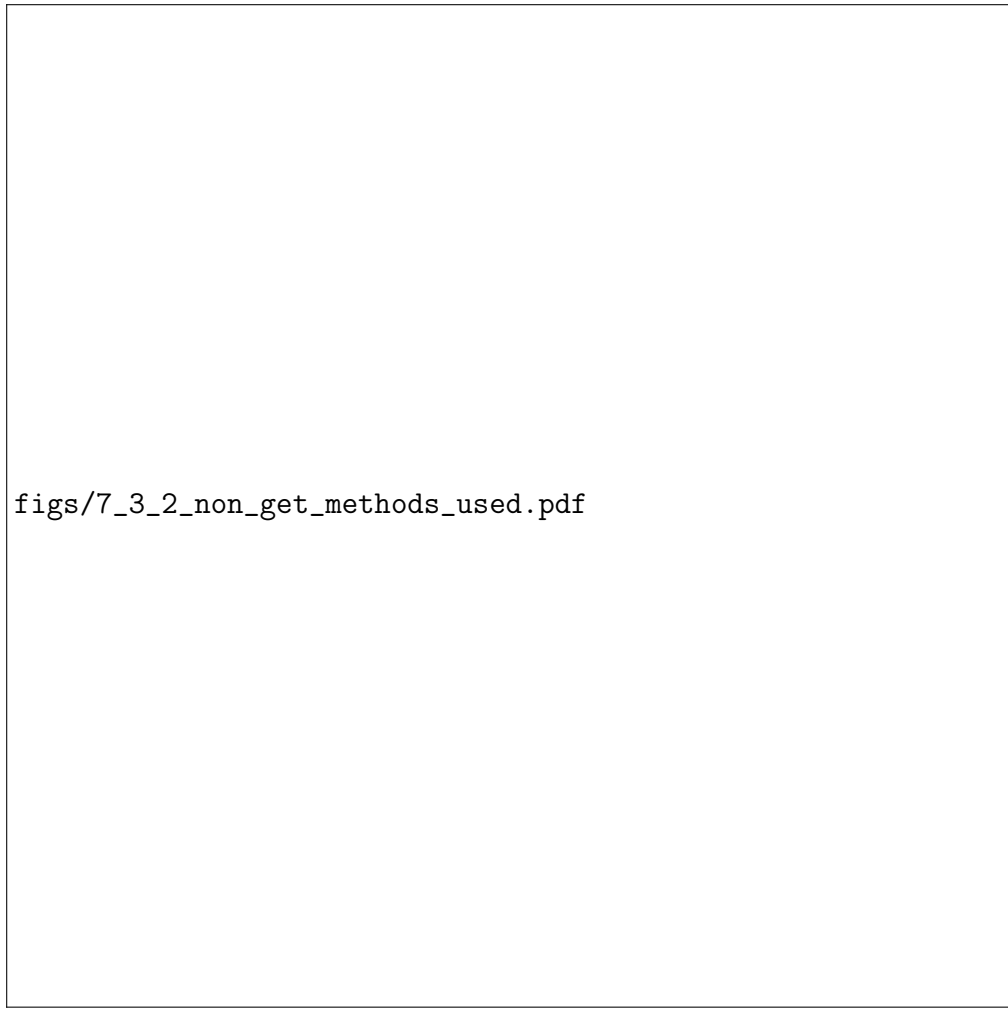
Figure 40: Proportions of resources requested by bots only



Figure 41: Proportions of query strings used in all requests with query strings



Figure 42: Proportions of HTTP methods used in each request




figs/7\_3\_2\_non\_get\_methods\_used.pdf

Figure 43: Proportions of HTTP non-GET methods used in each request

figs/7\_3\_3\_post\_trends.pdf

Figure 44: Aggregated hourly trends of HTTP requests with POST method



figs/8\_1\_1\_response\_codes.pdf

Figure 45: Proportions of all HTTP response codes



Figure 46: Proportions of all non-200 HTTP response codes

figs/8\_1\_3\_server\_errors.pdf

Figure 47: Number of aggregated server errors made per day in each month



figs/8\_1\_2\_forbidden\_attempts\_month.pdf

Figure 48: Number of aggregated *Forbidden* requests per day in each month

figs/8\_2\_4\_fastest\_response\_time\_by\_resource\_accessed.pdf

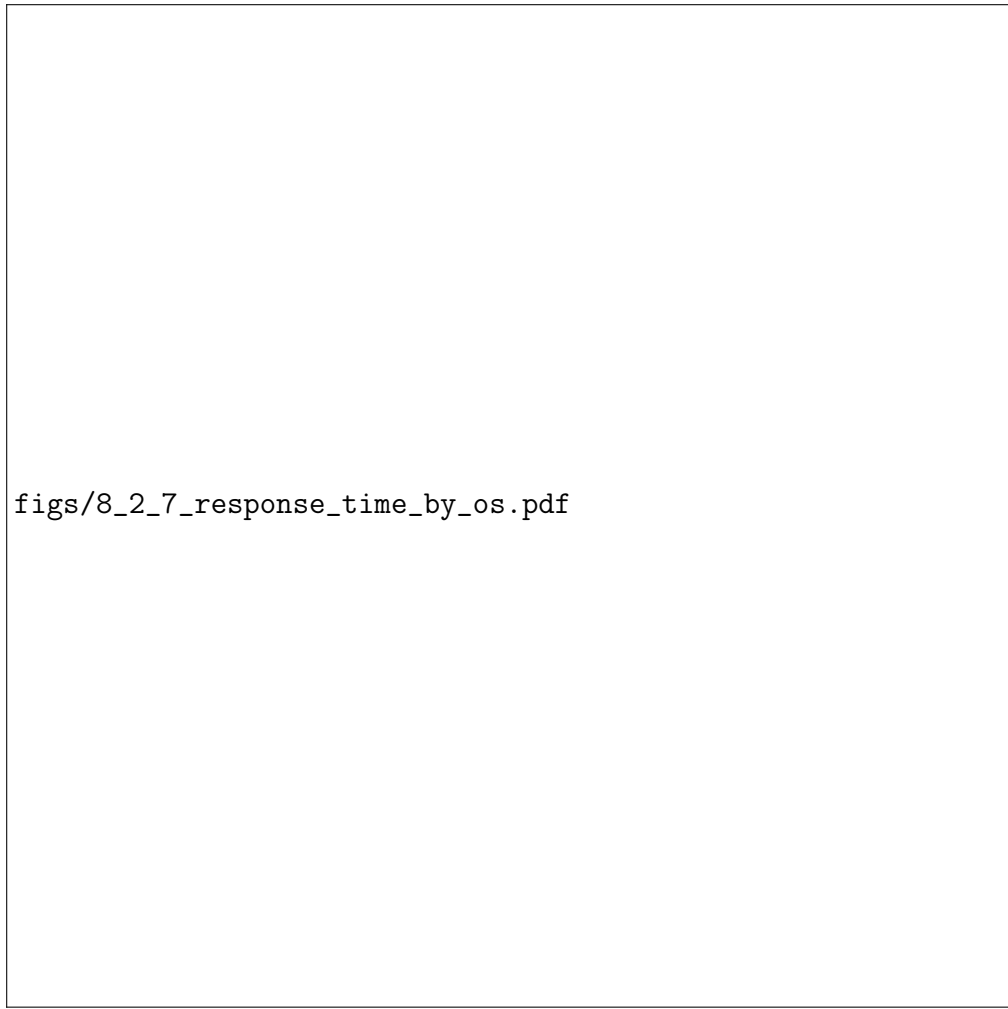
Figure 49: Resources with fastest average response times

figs/8\_2\_5\_slowest\_response\_time\_by\_resource.pdf

Figure 50: Resources with slowest average response times


figs/8\_2\_6\_top\_resource\_accessed\_response\_times.pdf

Figure 51: Top frequently accessed resources and their respective response times




figs/8\_2\_7\_response\_time\_by\_os.pdf

Figure 52: Average response time by operating system



figs/9\_1\_ports\_accessed.pdf

Figure 53: Number of times varying ports accessed



figs/9\_2\_https\_resource\_accessed.pdf

Figure 54: Most frequently accessed resources over HTTPS

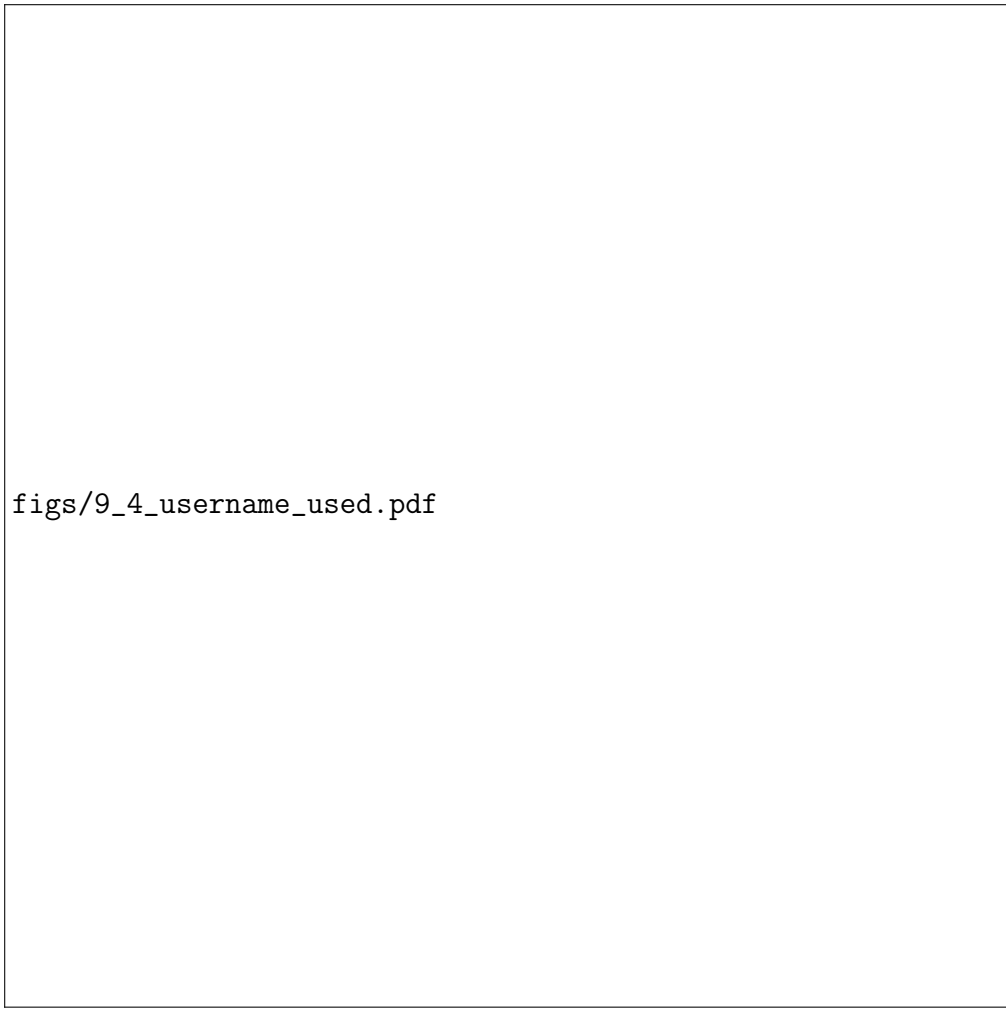
figs/9\_2\_https\_trends.pdf

Figure 55: Number of aggregated HTTPS requests per hour



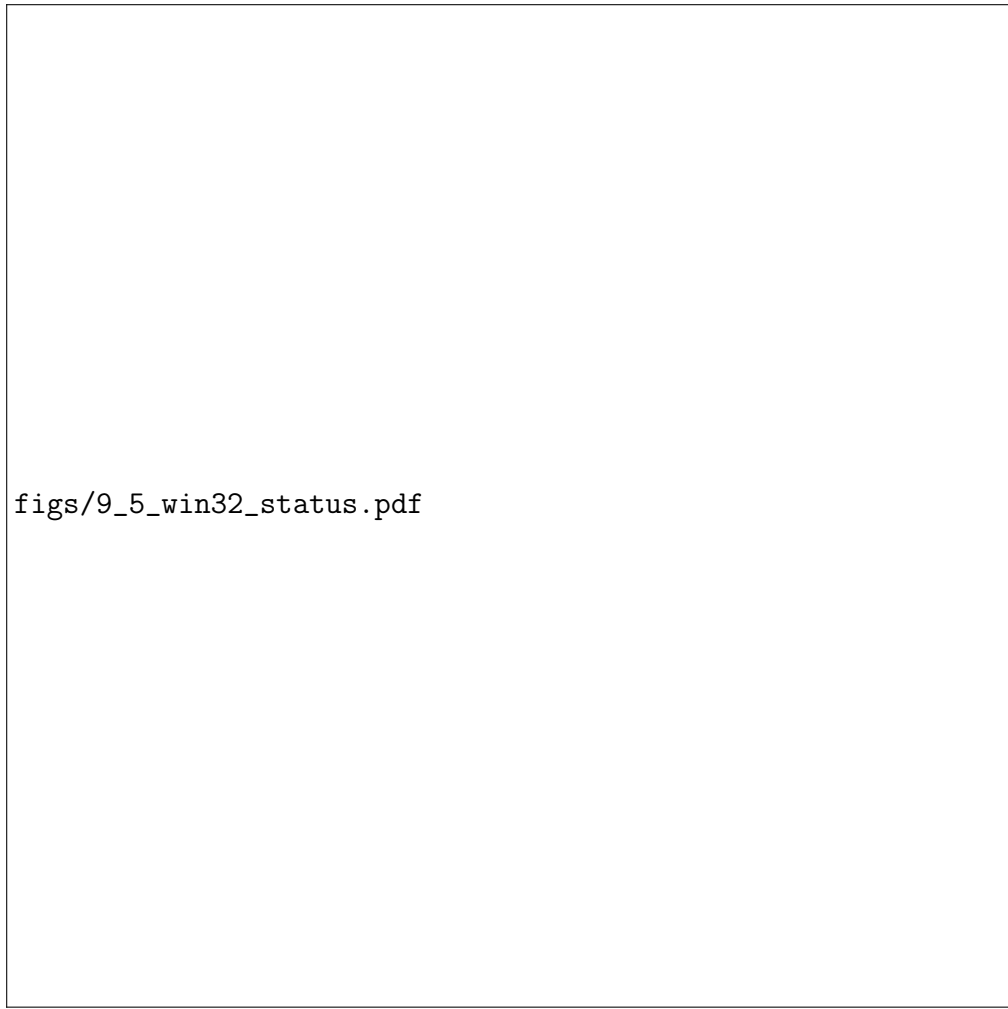


Figure 56: Number of aggregated HTTPS requests per hour



figs/9\_4\_username\_used.pdf

Figure 57: Proportion of non-empty usernames used

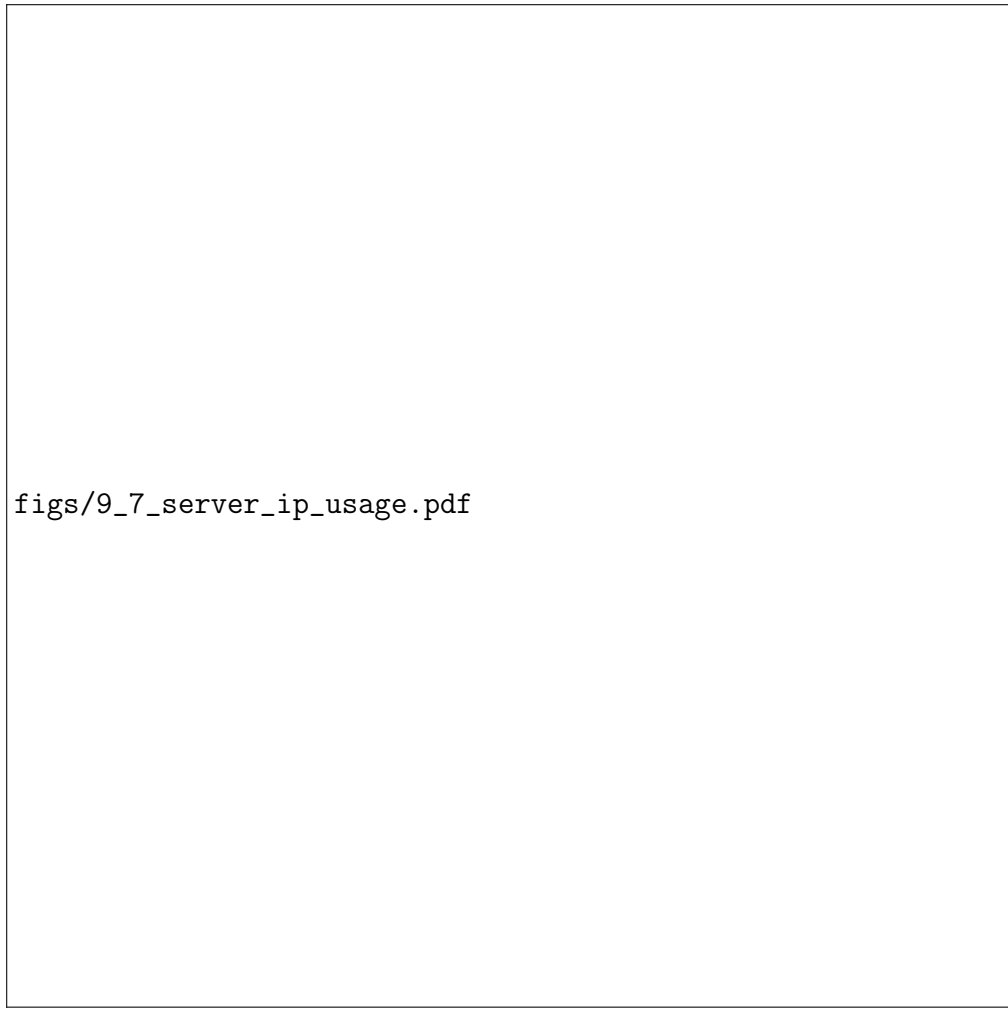


figs/9\_5\_win32\_status.pdf

Figure 58: Proportion of varying Win32 status codes

figs/9\_6\_iis\_substatus.pdf

Figure 59: Proportion of varying IIS substatus codes against HTTP status code



figs/9\_7\_server\_ip\_usage.pdf

Figure 60: Proportion of server IP usage

## **B   Extrapolation Results**

Attached on the following pages are the results from Databricks. You may also interact with this online on Databricks.

[fitpaper=true, pages=-]results.pdf