

Firefox Release - version (Fx 4.1)

1

LEAN &
Six Sigma

Service
Delivery
Committee

Stakeholders

Statistical
modelling

Cost Benefit
Analysis

New
Enhancement

UAT Testing



Full
Deployment



Firefox®

Our Success – User Experience

2

2004

Release of Firefox 1.0 and downloaded 100 million times in less than a year

December 2007

AOL announces end of Netscape and recommends users switch to Firefox.

January 2010

First release of Firefox on mobile device launched by Nokia

November 2010

Exciting pre-launch of Firefox version 4.1

The popularity of Firefox helps bring back choice to our

USERS



Firefox®



1(a). Business Problem & Hypotheses

3

Identify the Problem

Can we model the outcome of new user browsing experience using a number of tabs, to predict the factors that would influence a success launch for Firefox?

If yes, what are the primary variables that affect a user browsing experience?

Hypothesise goals



Null Hypothesis (H₀)

A 10% increase in the launch of new tabs will not significantly effect the browsing experience of Firefox users



Alternative Hypothesis (H₁)

We reject the null hypothesis and accept the alternative hypothesis that the launch of new tabs will have a significant effect on browsing experience for Firefox users.



1(b). Further Questions

4

Further questions for identifying the correct data set

- Survey Table

- How many of the total users completed the survey? **27,267 users**
- Of the users that completed the survey, identify the number of users who are new to Firefox and who are long time users **1819 users have been using Firefox. 215 are new users 'Less than 3 months' and 1604 are long time users 'More than 5 years'.** (Source: q1 of the survey).

- Bookmark Usage

- What's the median number of bookmarks? **29**
- What's the average number of bookmarks? **112.25**
- What fraction (%) launched a new bookmark? **99.97%**
- What fraction (%) of users, created new bookmarks? **20.11%**
- What is the distribution of how often bookmarks are used? Range, Mode, Max, Min
- **Range:19883, MIN: 0, MAX:19883, MODE:9**
- How does the number of bookmarks correlate with how long user has been using Firefox?

- Browser Tab Usage

- What is the distribution of maximum number of tabs? **MAX()**
 - Are there users who regularly have more than 10 tabs open? Open and > 10
 - What fraction (%) of user have ever had more than 5 tabs open? > 5
 - What fraction (%) of users have ever had more than 10 tabs open? > 10
 - What fraction (%) of users have had more than 15 tabs open? > 15

3. Identify and Obtain Dataset

5

Obtain the
data

+

Identify the
right data set



To investigate the current business problem, previous test data that was used in testing Firefox versions 3.5 ,3.6 and 4 Beta were collected from 527,817 test sets in November 2010 for Browser version 2.

The PostgreSQL hosted by Amazon Web included extraction of data from two servers:

1. Users & Survey tables

APAC: -analyticsga-apac.csuojbfcexhv.ap-southeast-2.rds.amazonaws.com

2. Events table:

US-West: analyticsga-cuwj8wvu6wbh.us-west-2.rds.amazonaws.com



The right test datasets were identified and stored in the Firefox database.

Table 1: Users – All the rows that have the same user number came from the submission. There are various ways that a single user could make **multiple submissions**

(E.g. multiple computers, multiple Firefox profiles). The order of user_id has been randomised. This is meta data of the users.

- Table 2: Events – Relates to the main table of users and recorded events which is defined by an event code.
- Table: 3: Survey – The table of users and associated Beta background survey answers comprised of the User_Id and 14 columns representing the survey questions.
 - The answers were in integers and coded in order of preference with '0' representing the first answer choice and '1' representing the second answer choice.
 - Users could select multiple answers
 - *Note : This survey was **optional** and did not include all users

4. Understand the data

6

Description of the data

To understand the data we read articles related to the Mozilla project community, the Mozilla timeline and also related information on desktop and mobile phone enhancements to the user browsing history to obtain background information

The three tables are related by the primary key 'User_id'.

Installed Add-On or Extensions:

MIN:1
MAX:185

27,267 users

7 columns

Data types (integer and text)

AVG:

940 Firefox versions

1514 OS

6816 test pilot extension version

*Note: includes duplication of user details with multiple submissions (multiple computers and Firefox profiles)

Events table

614,292 recorded events

7 columns

Data Type (integer, numeric and text)

NULL exists

Survey table

4081 surveyed users

15 columns

Data types (integer and text)

Missing values

Unreadable characters

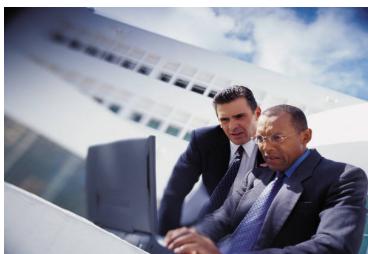


4.Understanding the data - Users Table

7

Who are our Users?

- Males and females who only use Firefox as their main browser
- and
- Males and females who use multiple browsers including Firefox, Opera, Safari, Chrome, Internet Explorer



Q. Describe the Users represented in the data. Do you think it's representative of the population of Firefox users ?

A. The Users in the data were not representative of the population of Firefox Users. The choice to complete the Survey was optional.



Summary of Users:

4081 users completed the survey

1819 users use Firefox (new= 215)

Less than 3 months', 1604 = long time 'More than 5 years'

1356 users only use Firefox as their daily browser (loyal customers) and,

2687 use other daily browsers:

551 (Chrome users)

139 (Safari users)

107 (Opera users)

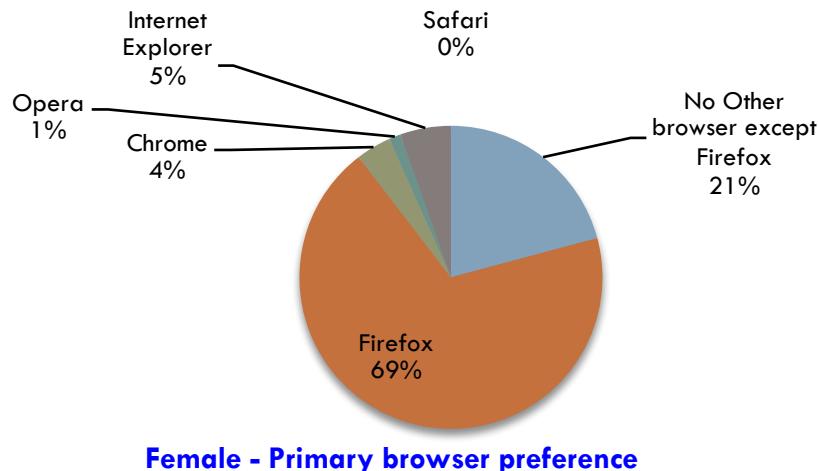
775 (Internet Explorer users)



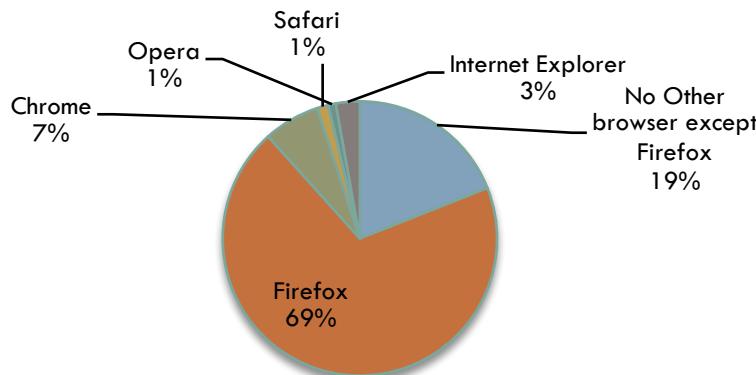
Firefox®

4. Understanding the data - Survey Table

8



Female - Primary browser preference



Male - Primary browser preference

Characteristics of Surveyed Users

The findings from the results of the surveyed users

- The primary browser preference of both males and females who participated in the survey favoured Firefox
- There were no females who chose Safari as their primary browser

Limitations of survey data

- The sample duration of 7 days was not long enough to provide an accurate prediction of trends as comparative time-series data of at least two years could not be collected.
- For q4, the female observation size was smaller compared to males (a total of which 3483 users were male and 240 were females)
- This was just a sample of data collected and the surveyed users do not represent the entire population of Firefox



Firefox®

5. Initial Draft Findings – Loyal Firefox users

9

Only 27,000 users were surveyed to test browser version 2 over a 7 day period, I do not think the number of users is representative of the entire population of Firefox users given that the test period was for a short duration.

Attributes of Users

- 4081 users completed the recent Beta survey. This survey was optional and conducted over 7 days
- From the respondents of the survey 1356 are our **loyal customer base** who only use Firefox as their only browser and >50% (i.e. 2687 users) use browsers from Chrome, Safari and Internet Explorer
- 1605 users have used Firefox for over 5 years and they are our long-term users and we also have 215 new users (i.e. Less than 3 months).
- Also we identified the number of users who use Firefox plus other browsers on a daily basis: 551 users use Chrome, 139 users use Safari, 107 users use Opera and 775 users use all versions of Internet Explorer. Surprisingly, Internet Explorer is our biggest competitor at the moment in terms of browsing experience.
- Profile for males and females who don't have other browsers but only use **Firefox** as their primary browser: returned 662 users were males and 50 were females who make Firefox their Primary browser.
- **Chrome Users:** males and females who use Chrome as their primary browser: includes 230 males and 9 females use Chrome as Primary browser
- **Safari Users:** Profile for males and females who use Safari as their primary browser included 47 males only who use Safari as Primary browser
- **Opera users:** 3 females and 31 males as their primary browser
- **Internet Explorer:** 98 males and 13 females make Explorer their primary browser.



6.Prepare, Structure and Clean the Data

10

- **SQL distinct** was not used for some calculations as we appreciated and wanted to reflect the usage behaviour of bookmarks where some users had multiple user devices and Firefox profiles
- **SQL cast** command was used to theoretically convert text strings into numeric values such as data1 variable under events table before performing aggregate functions but this generated syntax errors and hence data from event_code 8 was filtered using the WHERE BY clause and exported to excel for calculating aggregate functions such as AVG, MIN and MAX which must be numerical values only:
- `SELECT user_id ,data1 as TotalNumberofBookmarksfound ,data2 as TotalBookmarkFolders ,event_code as Bookmarkstatus ,COUNT(data1)FROM events WHERE event_code IN('8') GROUP BY event_code ,data1 ,data2 ,user_id ORDER BY data1 DESC;`
- **SGL Inner Join** command for was executed to join more than one table (i.e. Table events and survey) to obtain data from both tables by joining the two tables on a common primary-key the 'user_id' and also there no NULL values returned for this type of join.
- Run SQL queries with the **GROUP BY**(filtering the result set) and WHERE (limiting the output with conditions) clauses using the syntax with the event code of bookmarks to filter the results from the Tables events and survey. Results were sorted using **ORDER BY command** for further grouping



6.Prepare, Structure and Clean the Data

11

- Data was exported from SQL into Excel for the Bookmark launched analysis. The csv file was text to column and tab-delimited by comma into Excel.
- A new calculated field was used using IF logical statement to use the COUNTIF feature to calculate how many users launched at least one bookmark.
- Using conditional formatting to highlight the Top10 outliers in the column Bookmark launched
- SQL sub query
- Filters were placed on the column variables to examine the characteristics of the columns, check for null values and blank values.
- Summary statistics were entered to understand the mean, range, mode, median, max and minimum values of the result set
- Outliers were examined through a scatterplot in the preliminary exploration

5. Assumptions & Limitations

12

Assumptions

- We did not remove duplicate values from the User Id (primary key) because this unique identifier included users who had multiple Firefox profiles and could be browsing from multiple devices such as their smart phone and also their computers (iPad, laptop and desktop).
- There was no personally-identifiable information from the User table and this was anonymised to ensure data privacy.
- The 'user_id' is consistent across the three tables and hence the tables can be merged via the user_id.
- The user numbers have no correlation to anything outside
- All rows that have the same user number came from the same submission
- Each submission was a different user
- A single user can have multiple submissions (e.g multiple computers and multiple Firefox profiles)

Data Limitations

- We do not know if there were any mechanisms for control employed in the collection of data to smooth any data anomalies.
- The collection of test data is not a perfect representation of the entire Firefox population with approximately 27,000 users but represents a sample of data collected at a point in time.
- The data collected from the web included missing values, free form text and non-readable characters in the survey data which required cleaning and transform before the data could be used for data analysis or prepared for a machine-readable state.
- There was not enough data collected to draw conclusions or inferences because historical time-series data was not collected for comparative analysis.
- It was not feasible to compute a histogram for Bookmarks usage because the range of bins was too large
- For data optimisation in data extraction from the database SQL, the query of event records was only limited to a sample observation size of 27627 which is equivalent to the number of users.



5. Bias

Bias

- The collection of data was for a duration of 7 days and this short time frame represents user activity over a point in time and therefore bias is present in the sampled data.
- ‘Time will tell data’ bias was that not enough data was collected over the project delivery time and this would effect the predicted outcomes produced in the results .
- **Sampling bias** -Again, the sampling size of 27,000 users does not represent the entire population of Firefox users because the survey was not completed by the whole population like a census but it was optional and users had choice to opt-out.
- **Sub-group sampling bias-** The initial findings when exploring data gave us insight that the sampled users who completed the survey were highly skewed for male participants. For example when breaking down the sub-group of
- **Modelling bias** -

Exploratory Analysis – Bookmark usage

What's the median number of bookmarks? 29

What's the average number of bookmarks? 123.15

What fraction (%) of users, launched at least one bookmark during sample week? 99.97%

What fraction of users created new bookmarks?

What is the distribution of how often bookmarks are used?

Mode =9

Min =0

Max =19833

Range =19833

Standard Deviation(sample) =509.44

How does the number of bookmarks correlate with how long user has been using Firefox? Sub query
(inner query in main query)



Exploratory Analysis – Browser Tabs

- Summarize the exploration of usage of Brower tabs in the presentation:
 - What is the distribution of maximum number of tabs? **89823**
 - Are there users who regularly have more than 10 tabs open?
 - What fraction (%) of user have ever had more than 5 tabs open? >5
 - What fraction (%) of users have ever had more than 10 tabs open? > 10
 - What fraction (%) of users have had more than 15 tabs open? >15
- Description of browser tabs
 - Sample size : **89823** (some users have multiple profiles)
 - MIN:0
 - MAX: **89823**
 - RANGE:**89823**
 - Median:**3**
 - Mode:**2**
 - Missing values/Blank values: **NIL**



8. Analyse Results & Interesting Insights

16

- By removing the Top 10 outliers in the variable bookmark launched (data1) in event_code 8, using a scatterplot the distribution of data points was still non-linear in shape and highlighted more outliers as they were numerous. So we decided to retain the outliers for this analysis and we understand the mean, mode, range and min and maximum values will be skewed.

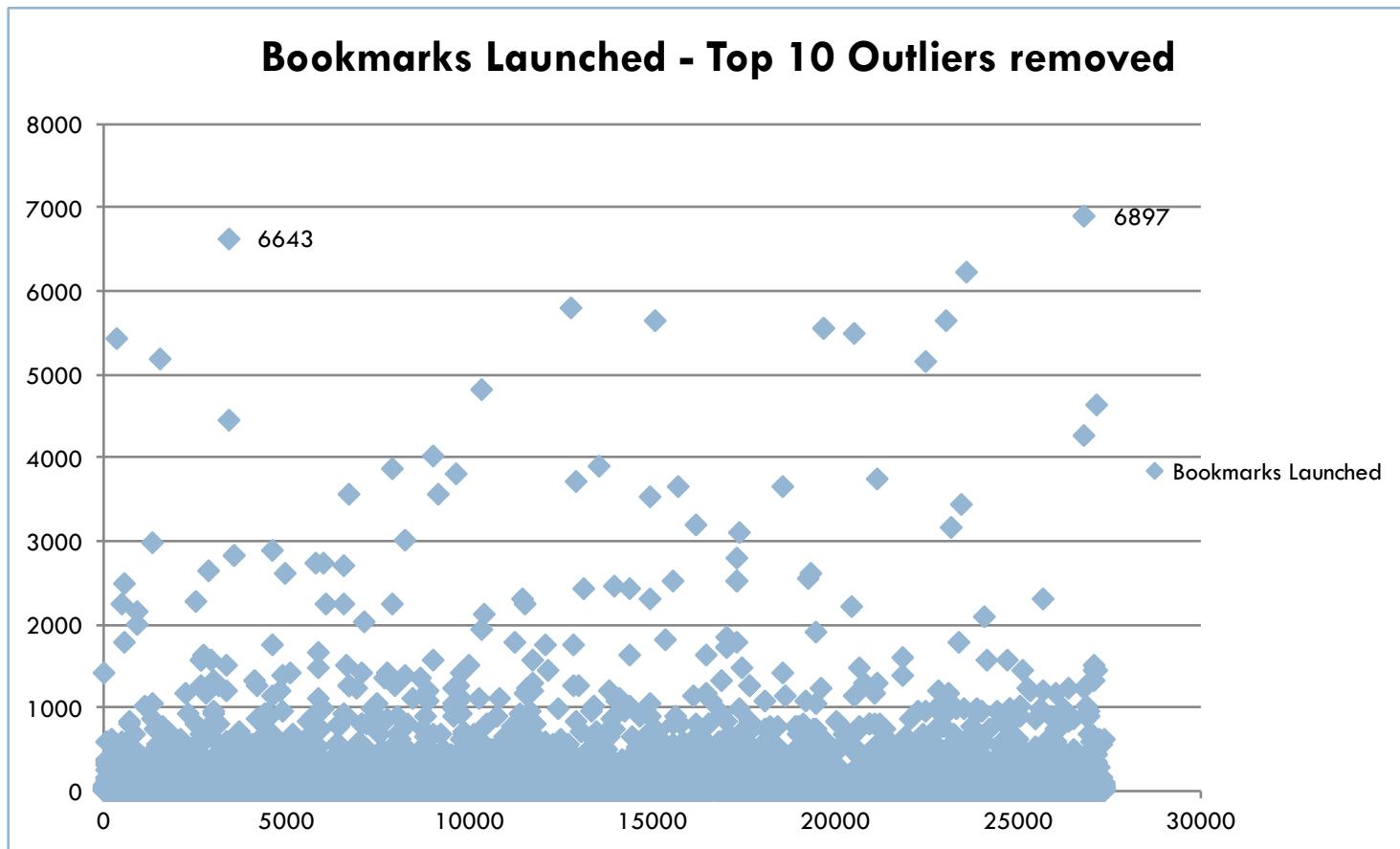
8. Analyse Results & Interesting Insights

17

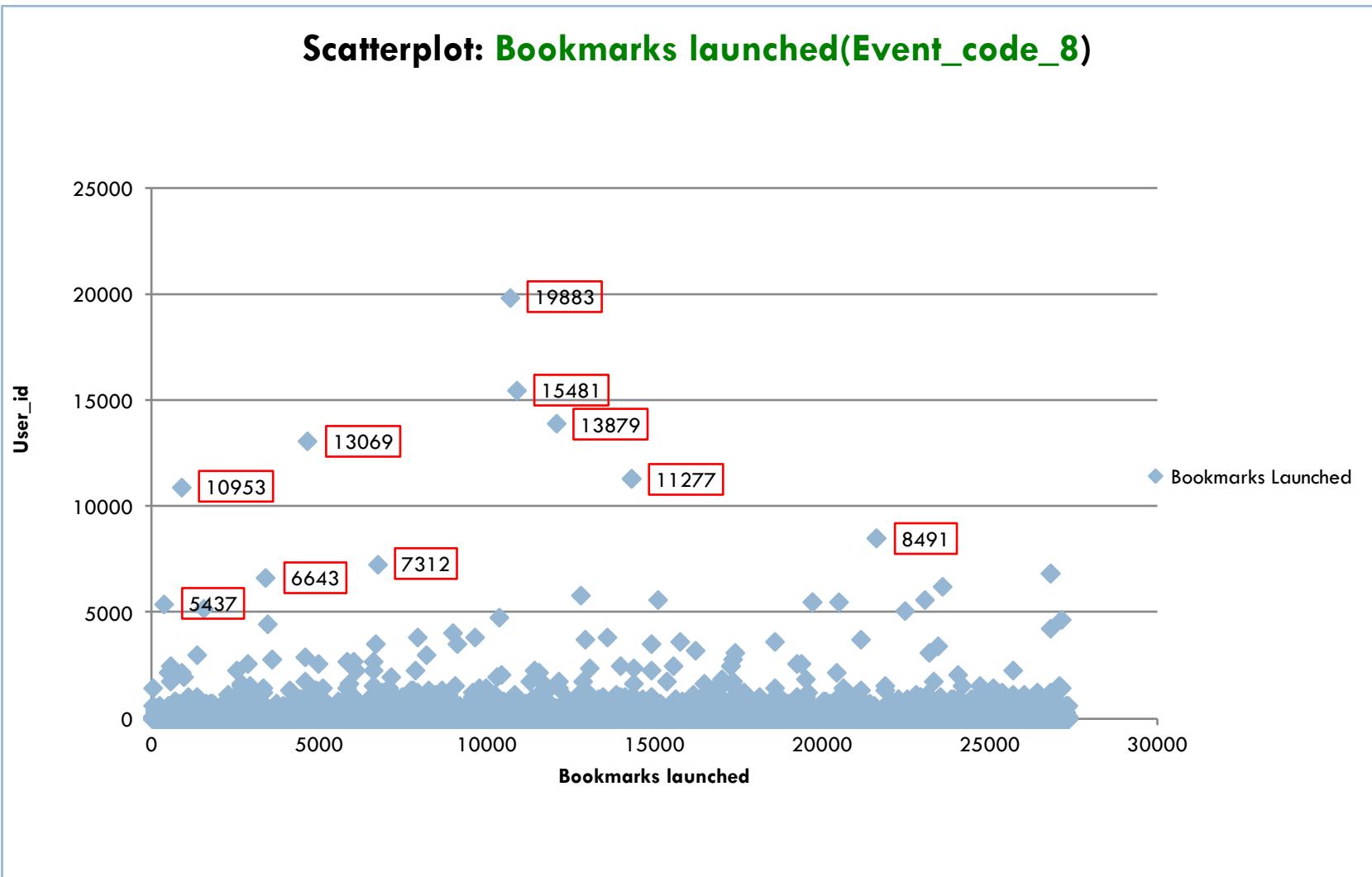
- By removing the Top 10 outliers in the variable bookmark launched (data1) in event_code 8, using a scatterplot the distribution of data points was still non-linear in shape and highlighted more outliers as they were numerous. So we decided to retain the outliers for this analysis and we understand the mean, mode, range and min and maximum values will be skewed.

8. Analyse Results & Interesting Insights

18



Outlier Analysis – Bookmark usage



9.(b) Outliers – Data Analysis

20

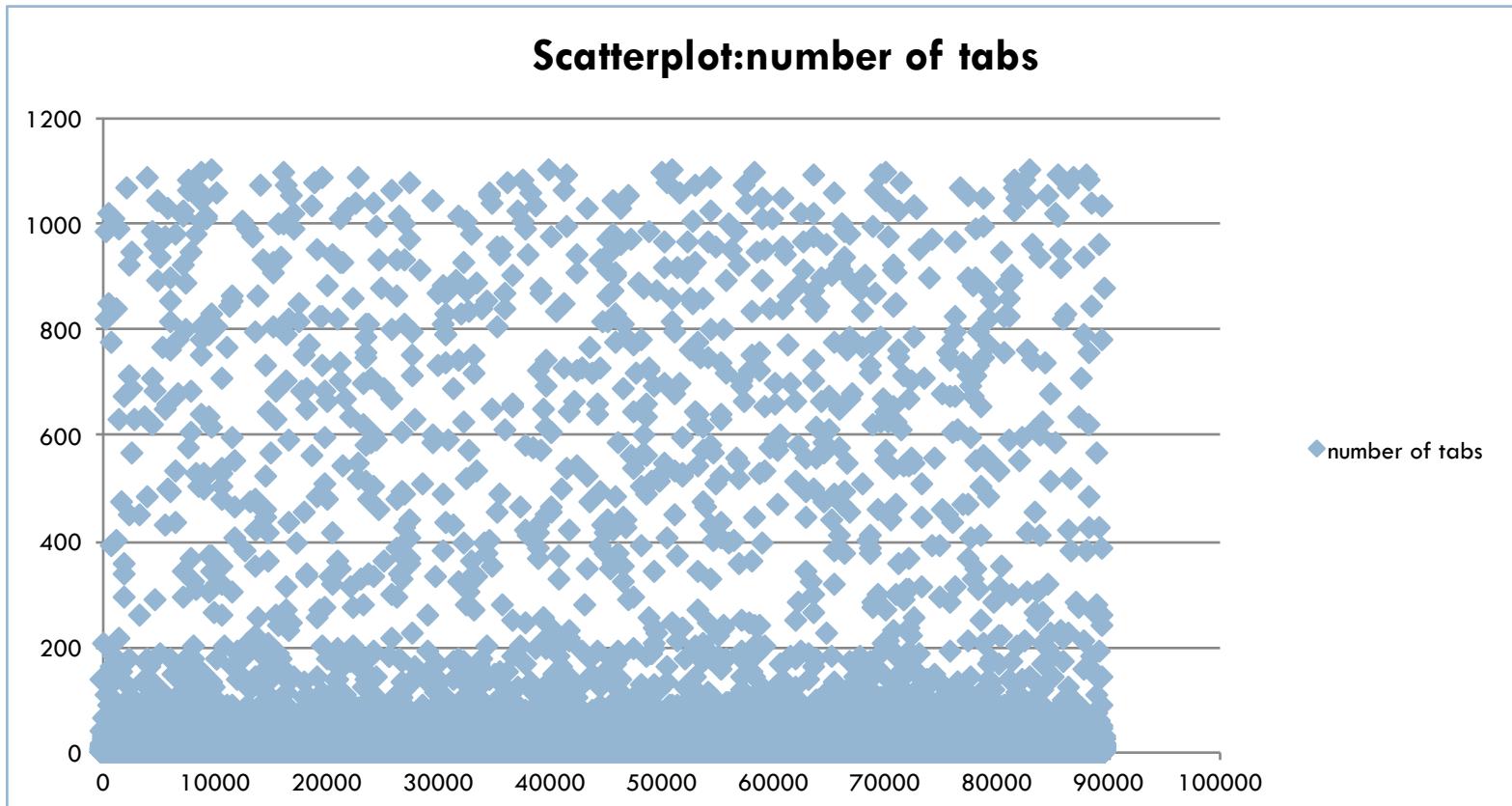
- Scatterplot helps us detect the non-linear fit of Bookmarks launched. Outliers at top 10 observations

Top 10 Bookmarks Launched (Outliers)
7312
7312
8490
8491
10953
11276
11277
13069
13879
15481
19883

- Remove or retain outlier from the factor variable
- We did not remove the outliers because

Outlier Analysis – Browser Tabs

21



5. Missing Values (NULL Values)

22

NULL values (“N/A”)

- The variable Bookmarks launched under event_code8 did not have any missing values (“N/A”) but recorded valid zero results for some users who did not launch any bookmarks upon the start of their browsing session.

8. Multiple linear regression

23

- Assumptions: for multiple linear regression
- Explanatory variables –why certain variables selected in the model
- Collinearity among variables

- Limitations
- Intercept
- Residuals
- P-value
- Adjusted R-square
- Histogram, Scatterplot

8. Multiple linear regression

24

- Assumptions: for multiple linear regression
- Explanatory variables –why certain variables selected in the model
- Collinearity among variables

- Limitations
- Intercept
- Residuals
- P-value
- Adjusted R-square
- Histogram, Scatterplot

8. Statistics – Browser Tabs

25

- MIN
- MAX
- MODE
- RANGE
- STANDARD DEVIATION(sample)
- Variance
- Mean
- Histogram, Scatterplot

8. Statistics - Bookmarks

26

- MIN
- MAX
- MODE
- RANGE
- STANDARD DEVIATION(sample)
- Variance
- Mean
- Histogram, Scatterplot

Correlation Test – Two-Paired Test (2 factors)

27

- Pearson Correlation Coefficient
- Scatterplot of two explanatory variables
- Scatterplot of two explanatory variables
- Correlation does not mean causation

Box Plot

28

- Check for normal distribution - factor variables
- Check outliers – retain or remove outliers from the variable?

8. Statistics - Bookmarks

29

- MIN
- MAX
- MODE
- RANGE
- STANDARD DEVIATION(sample)
- Variance
- Mean
- Histogram, Scatterplot

8. Statistics – Bookmarks created Top 10 outliers removed

30

- MIN
- MAX
- MODE
- RANGE
- STANDARD DEVIATION(sample)
- Variance
- Mean
- Histogram, Box plot



8. Statistics - Tabs

31

- MIN
- MAX
- MODE
- RANGE
- STANDARD DEVIATION(sample)
- Variance
- Mean
- Histogram, Box plot

8. Analyse Results & Interesting Insights

32

- Present the results of the analysis – based on your findings, create a coherent argument (bookmarks or tabs?)
- Loyal customers
- Users likely to switch (profile)

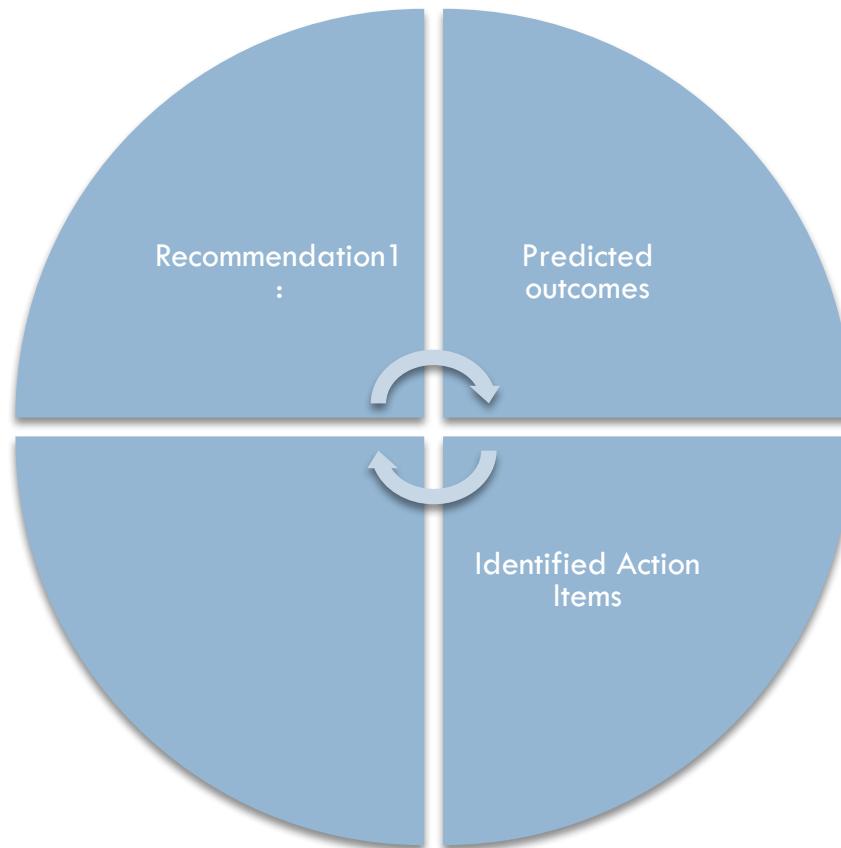
10. Presenting Results – Analyze data and Visualisation

33

- Dashboards and Visualisation –scatterplot, bar chart, customer segments
- To determine findings and create a recommendation
- Pie charts, linear trends
- **Consumer segmentation**
- Validation of data
- Simple metrics – top 10%, rank
- Conditional formatting
- Auto filters
- Export data to Excel to create any visualisations to help explore data or provide evidence for your recommendation

11. Recommendation: User segmentation

34



Recommendation— Bookmarks vs Tabs

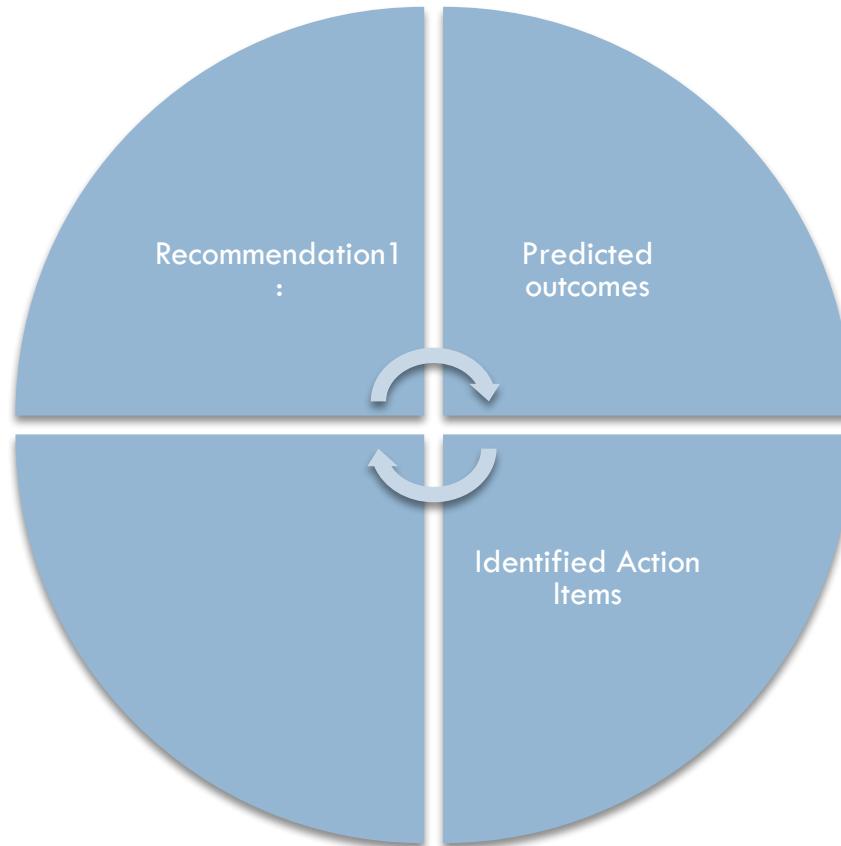
- Compare the two features and create a coherent recommendation (to your boss)
 - Do you agree with opinionated colleague?
 - Do you think the team should go in a different direction?

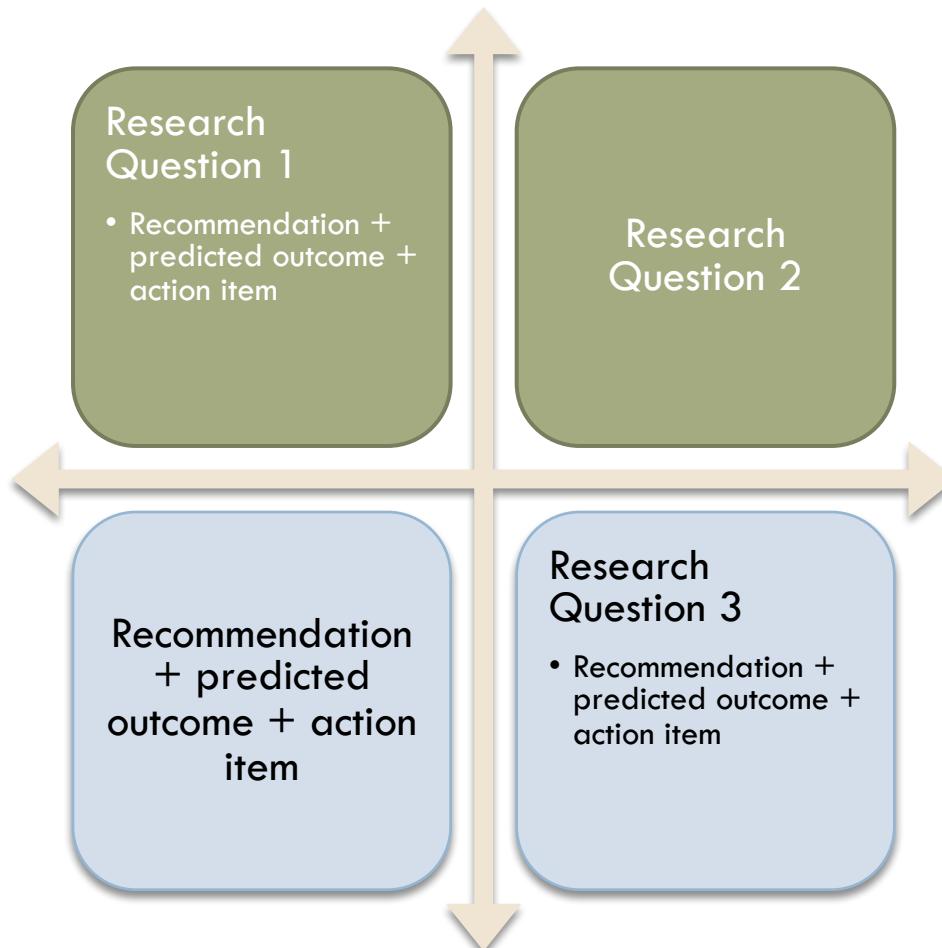
Recommendation— Bookmarks vs Tabs

- Compare the two features and create a coherent recommendation (to your boss)
 - Do you agree with opinionated colleague?
 - Do you think the team should go in a different direction?

11. Recommendation: Firefox Market share

37





12. Summary of Recommendations

39

13. Follow-Up Problems - Further Analysis

40

Q1. How do we target our marketing efforts to users who are likely to switch to our competitors?

Q2. What is the marketing budget to attract new clients to Firefox?

Q3. When and how do we collect more data over time for comparative analysis

Q5. Can make better predictions using alternative models instead of multiple regression

Q6. Is there a budgeted expenditure for a dedicated resource for the remediation of new tabs if any issues are identified?

Q6. Is there a budgeted expenditure for a dedicated resource for the remediation of new tabs if any issues are identified?

14. What are our Next Steps?

41

Action Items:

Ensure that we have received final legal sign off by Compliance and IT departments

Ensure that all completed UAT testing scripts from selected users have been received and returned to the Test team

Ensure the process is double -checked before go-live production

Delivery Timeframe

December 2010 – Business deployment of Firefox version 3.5 over Christmas shut-down period so that the IT department have full access and minimal disruption for full deployment during business hours.

Next Meeting

January 2011 – The Service Delivery Committee will reconvene over the new year break and measure and evaluate the success of the deployment and the marketing team will provide key metrics via web analytics to test the adoption rate of new tabs by existing or new users to Firefox.



Conclusion

42

Thank you
Q & A?

