METRICS OF SUCCESSFUL WEBSITES AND COMPANIES

Danai Avratoglou | Thesis Presentation | Master in Business Analytics

Contents

Introduction

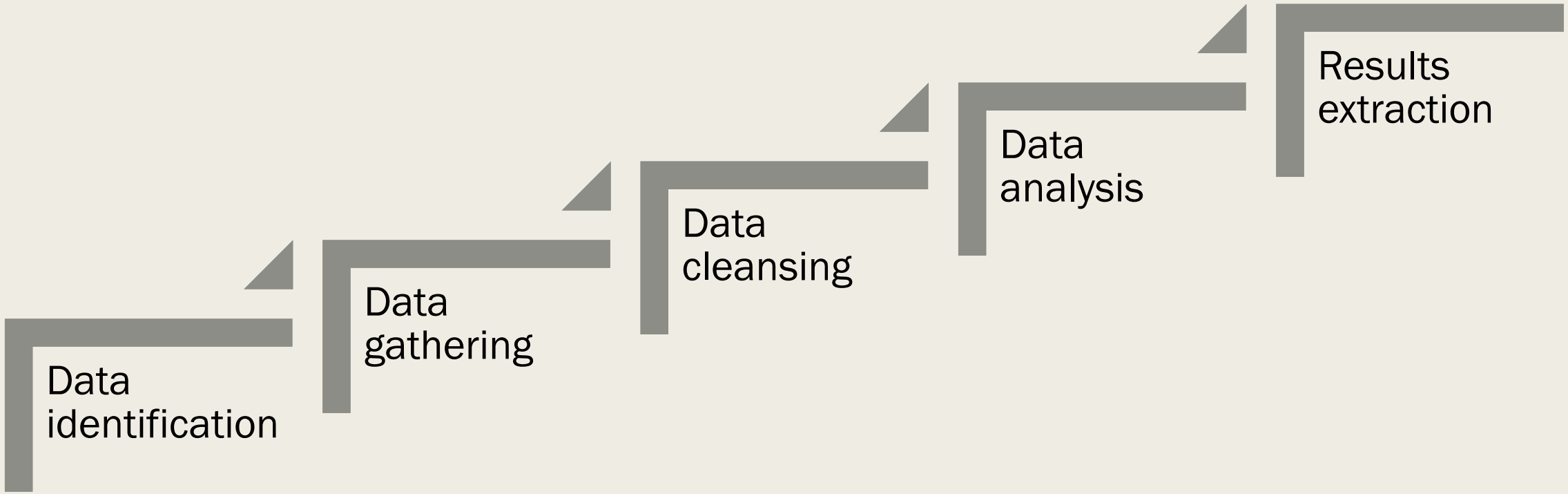
Research Goals

- ❑ Understand the correlation between a website’s specific metrics and a company’s status/success
- ❑ Test metrics’ categories that were deemed important by user’s and web designer perspectives
- ❑ Evaluate the Fortune 500 companies but without any industry deviation

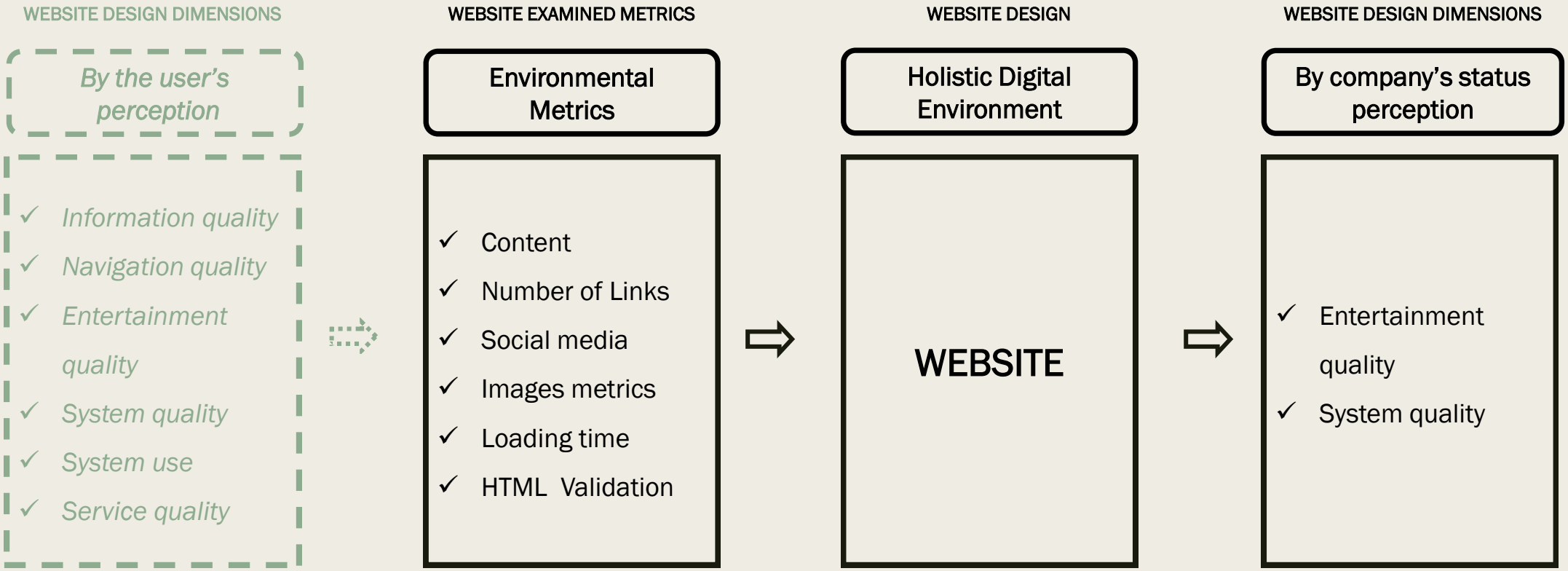
Research Hypothesis

- ➢ Do the revenues of a company correlate to specific metrics (or to some of the 6 dimensions of a website's design quality based on the user's perception) of the company's website?
- ➢ Which of the metrics under examination are correlated (individually) the most with the revenues of each company of the Fortune 500 ones?
- ➢ Which of the metrics under examination are correlated (in groups) the most with the revenues of each company of the Fortune 500 ones?

Research Framework



Data identification / Research model



Literature Review

Previous Researches

1997	• Web sites of the Fortune 500 companies: Facing consumers through home pages, Liu C. • Content analysis of website features • 2/3 of the companies had already websites
1999	• Differences in Public Websites: The current state of Large U.S. Firms • Only 10 companies still without a website
2000	• Exploring the factors associated with Web site success in the context of electronic commerce, Liu C.
2005	• Analysis of critical websites characteristics : A cross category study of successful websites • User's and web designer's perspective
2006	• Web site practices : A comparison between the top 1000 companies in the US and Taiwan, Liao C.
2010	• Factors driving website success : the key role of Internet customization and the influence of website design quality and Internet marketing strategy, Wei-Shang F.

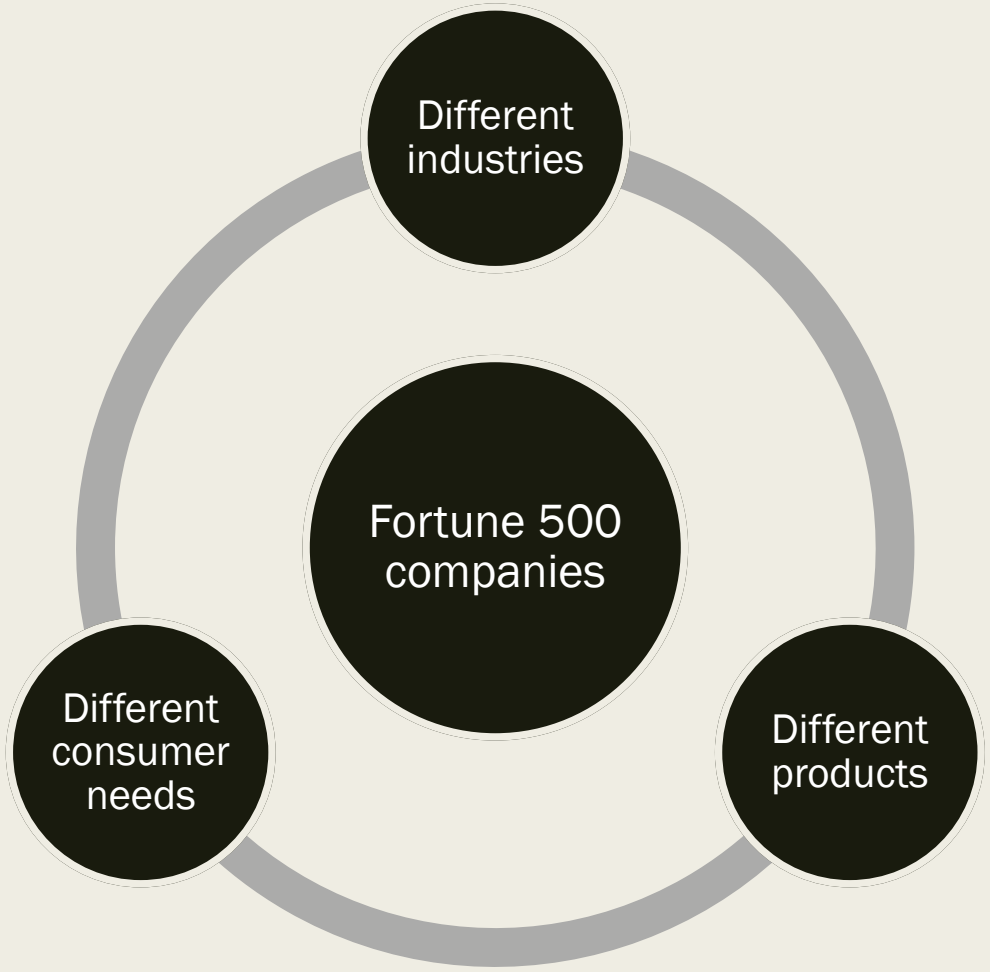
Research Gap

- *Study the most successful companies, without separating them in industries*

- *Study website metrics not from a user's or a web designer's perspective but from the actual companies status*

Data Gathering

Fortune 500 companies (2016)



- ✓ Assets
- ✓ Ranking
- ✓ Market Value
- ✓ Total Stock holders Equity
- ✓ Revenues
- ✓ Profit % Revenues

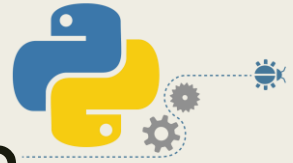
Metrics Categories

- Loading time
- Hyperlinks
- Social media
- Image metrics
- Content
- Html Validation

Website Metrics

- ✓ Loading time
- ✓ External, Internal, and total links
- ✓ Facebook, Twitter, Instagram, Pinterest, YouTube, LinkedIn
- ✓ Pixels, Formats, Total
- ✓ Readability Index, Words, Unique words, Sentences
- ✓ Number of errors, number of warnings

Data collection tool: Python



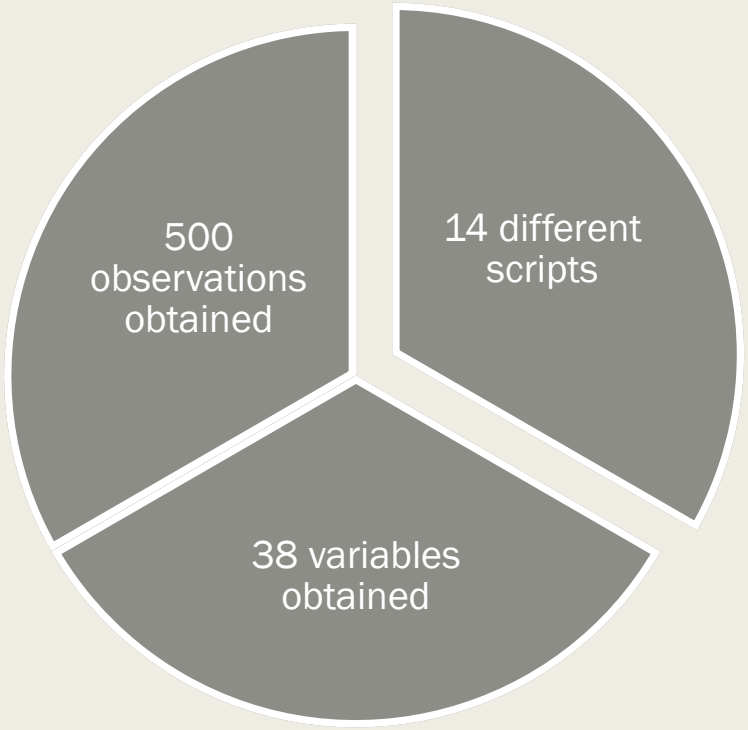
- High-level programming language
- Convenient syntax
- Easily readable

Python 2.7 version

jupyter

ANACONDA

Python scripts



Crawling websites

- Create a fake browser
- Open the page and downloaded with urllib2
- Use time library to leave time space between different openings

Keeping specific data

- regex
- soup.HTML.body.findAll()
- split
- re.findall

Libraries

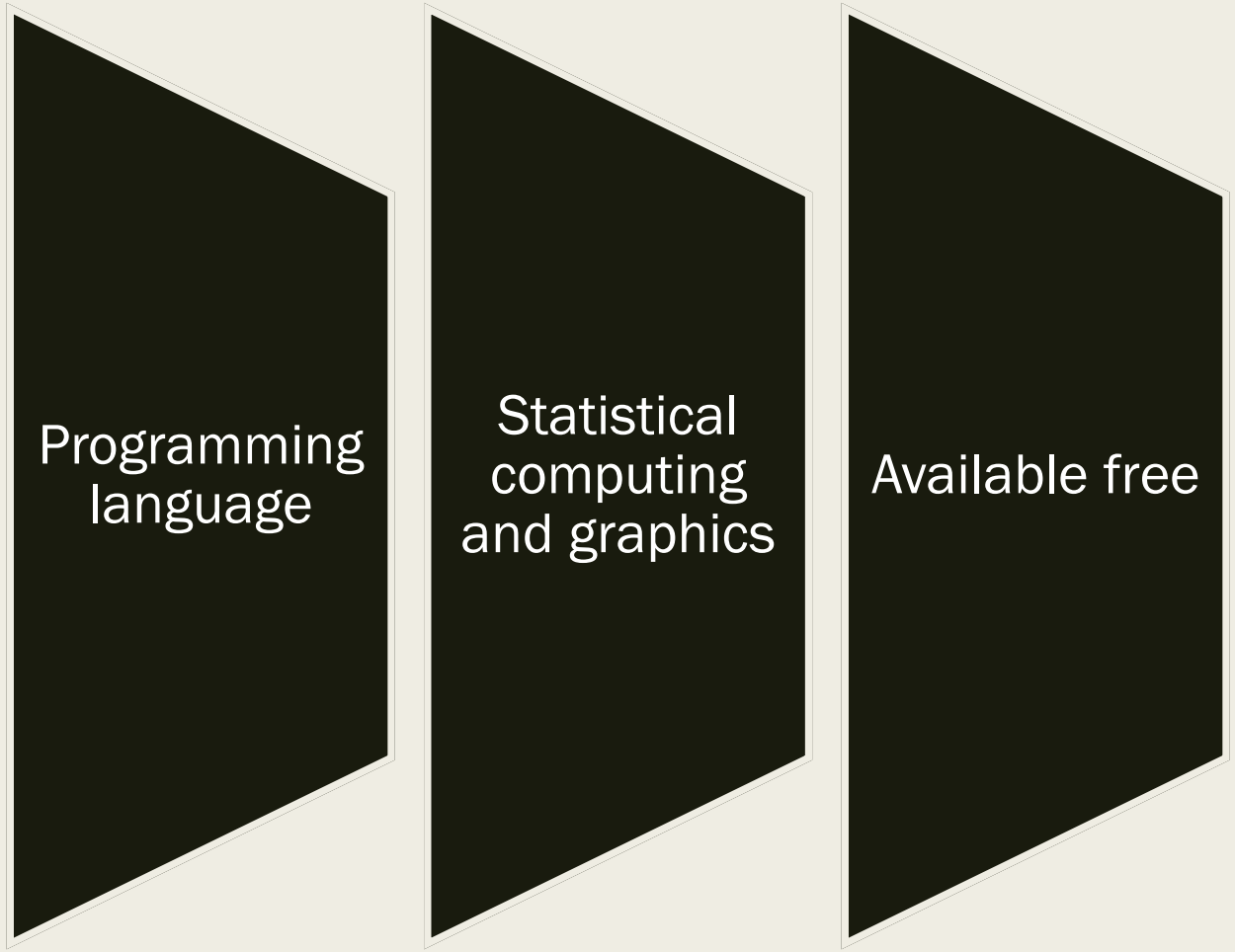
- pandas
- beautiful soup
- time
- urllib2
- re
- Validators

Other websites used

- Html validator
- Readability tool
- Fortune 500 list from article

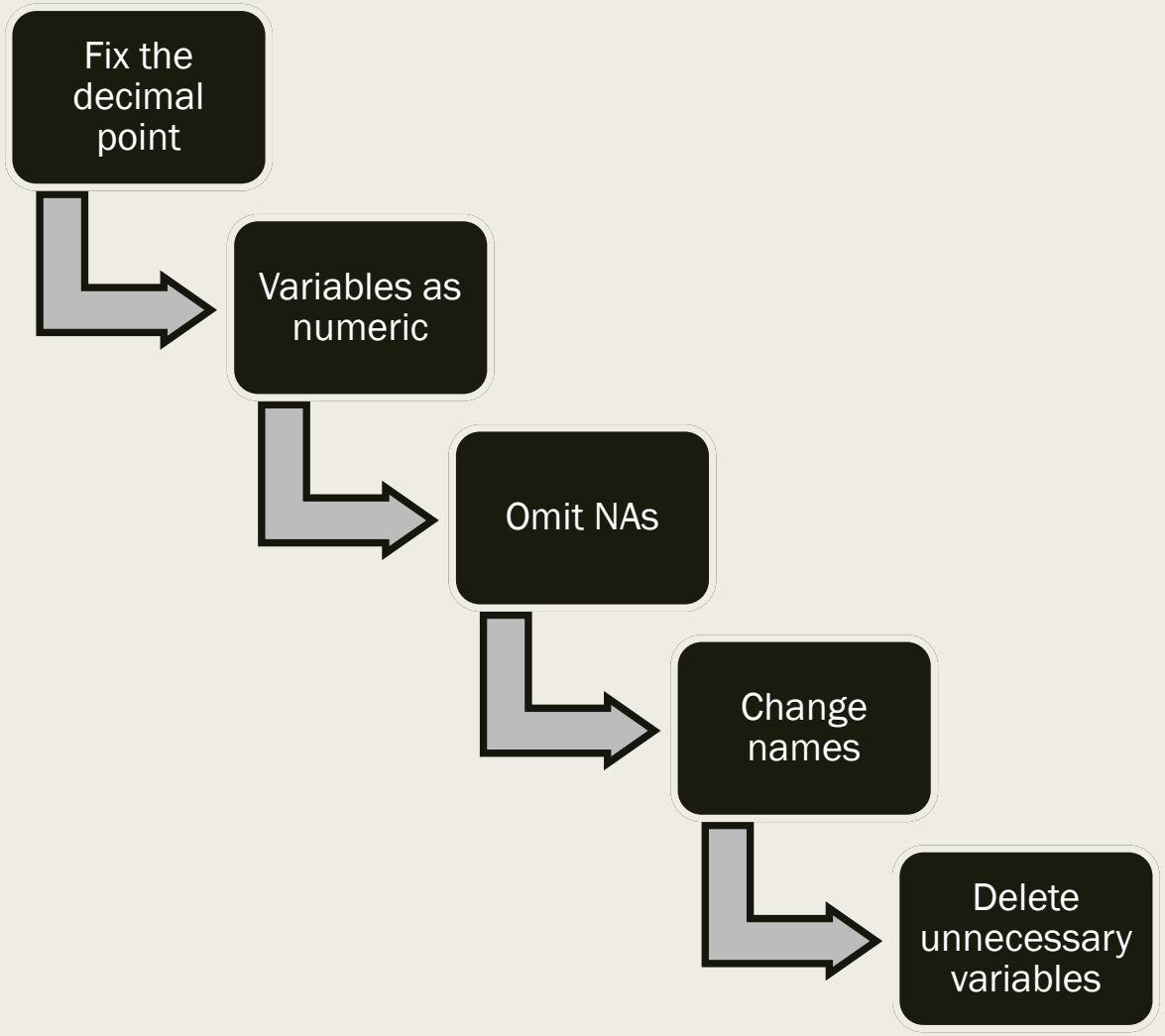
Data Analysis

Data analysis tool: R



RStudio®

Data Cleansing



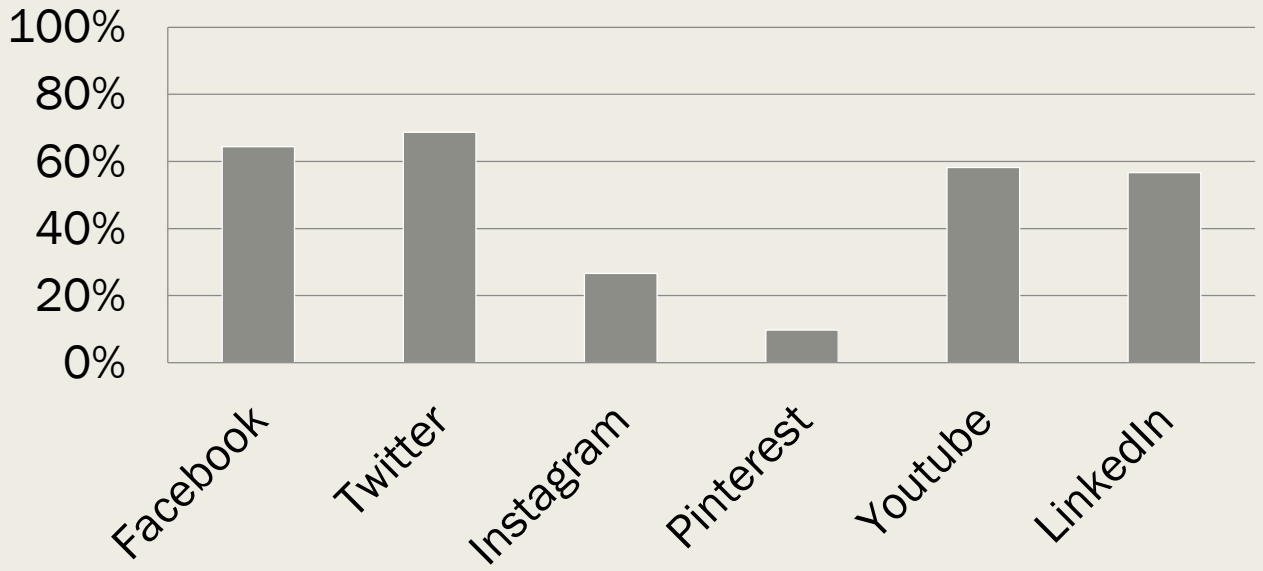
Dataset	Observations	Variables
Initial dataset	500	38
After cleansing	402	36

Fortune variables analysis

Summary	Ranking	Assets	Market Value	Total Stockholder equity	Revenues	Profit % Revenues
Min.	2	1,077	1,022	-978	5,13	-347,4
1st Qu.	134,5	7,72	6,93	2,848	7,114	2,3
Median	254,5	17,588	16,294	6,545	11,234	6,05
Mean	252,7	51,152	49,435	66,395	22,34	5,693
3rd Q.	370,8	42,224	39,797	20,469	20,867	11,7
Max.	499	877,933	947	998	233,715	45,6

Regression model	Adjusted R-squared	p-value
model_ranking_revenues	0,4531	< 0,05
model_ranking_Profit_per_Revenue	-8,32E-02	0,3261
model_ranking_assets	0,1351	1,57E-11
model_ranking_market	0,0226	0,001458
model_ranking_equity	0,01322	0,01198
model_fortune500_group	0,4701	<0,05

Social media variables analysis



Regression model	Adjusted R-squared	p-value
model_revenue_facebook	-0,0024	0,829
model_revenue_twitter	-0,0025	0,992
model_revenue_instagram	0,0007	0,260
model_revenue_pinterest	0,0079	0,041
model_revenue_youtube	0,0028	0,145
model_revenue_linkedin	-0,0024	0,866
model_revenue_socialmedia	0,002	0,336

Links variables analysis

Summary	External links	Internal links	Total links
Min.	0	0	0
1st Qu.	2	73	81
Median	5	116,5	133
Mean	17,83	154,2	172
3rd Q.	13	182,8	203,8
Max.	545	1254	1255

Regression model	Adjusted R-squared	p-value
model_revenue_totallinks	-0,002	0,753
model_revenue_external	-0,001	0,503
model_revenue_internal	-0,002	0,918
model_revenue_links	-0,004	0,795

Content and loading time variables analysis

Summary	Sentences	Words	Unique Words	Flesh measure	Readability index
Min.	1	1	0	-3422,4	1
1st Qu.	70,5	285,5	59	34,62	1
Median	139	500	109	45,55	3
Mean	176,6	685,8	151,1	35,22	2,94
3rd Q.	239,8	894,2	187,8	55,4	5
Max.	1350	8306	1910	121,2	7

Regression model	Adjusted R-squared	p-value
model_revenue_loading_time	0,013	0,014
model_revenue_Sentences	-0,002	0,689
model_revenue_Unique,words	-0,001	0,403
model_revenue_Words	-0,001	0,452
model_revenue_Flesh_Mesaure	-0,002	0,619
model_revenue_Readability	-0,002	0,728
model_revenue_r	0,001	0,360

Html Validation variables analysis

Summary	Non document error	Number of errors	Number of warnings
Min.	0	0	0
1st Qu.	0	0	0
Median	0	13	3
Mean	0,2264	37,49	8,684
3rd Q.	0	37	9
Max.	1	995	214

Regression model	Adjusted R-squared	p-value
model_revenue_number_of_errors	0,004	0,109
model_revenue_number_of_warning	0,007	0,053
model_revenue_non_document_error	0,003	0,138
model_revenue_html	0,007	0,118

Image variables analysis

Summary	jpg	png	tif	tiff	total images	pixels
Min.	0	0	0	0	0	13.220
1st Qu.	2	3	0	0	13	2.148.967
Median	8	8	0	0	24	4.048.858
Mean	18,12	15,46	4,21	0,014	47,96	5.804.762
3rd Q.	15,75	17,75	3	0	42,75	6.750.073
Max.	363	304	301	2	2162	55.476.065

Regression model	Adjusted R-squared	p-value
model revenue total images	0,000	0,306
model revenue pixels	-0,002	0,712
model revenue bmp	0,005	0,094
model revenue dib	0,005	0,079
model revenue gif	-0,002	0,651
model revenue jpe	0,001	0,233
model revenue jpeg	0,001	0,234
model revenue jpg	-0,002	0,921
model revenue png	-0,002	0,672
model revenue tif	-0,002	0,943
model revenue tiff	-0,002	0,541
model_revenue_im	0,001	0,419

Multiple Regression analysis

Regression model	Adjusted R-squared	p-value
full_Revenues	0,005	0,361
full_Ranking	0,017	0,181
full_Profit_per_Revenue	-0,045	0,999
full_Market_Value	0,034	0,048
full_Assets	0,014	0,218
full_Total_SH_Equity	-0,016	0,800

Findings

Conclusions

➢ Do the revenues of a company correlate to specific metrics (or to some of the 6 dimensions of a website's design quality based on the user's perception) of the company's website?	***Small correlation with two dimensions the entertainment quality and the system quality***
➢ Which of the metrics under examination are correlated (individually) the most with the revenues of each company of the Fortune 500 ones?	***Loading time, formats (bmp, dib), number of warnings***
➢ Which of the metrics under examination are correlated (in groups) the most with the revenues of each company of the Fortune 500 ones?	***Loading time, formats(bmp, dib) In total group no correlation in the subgroups***

Theoretical contribution

- Confirmation of the website quality dimensions from the site of website status
- Correlation aside of industries
- Not user's oriented results

Managerial contribution

- Companies - help in website construction
- Test/apply the findings – Improve the results
- Web designers – apply the findings – improve quality

Further Research

Further Research Opportunities

- Test a larger number of companies
- Test in different countries
- Use metrics given from the companies

THANK YOU FOR YOUR ATTENTION

Any questions?