**Background**

Search Engines are the dominant method people use to find information today. Around 82% of all internet users use search to navigate online. For many companies ranking well in search engines is business critical and a large marketing budget is spent each year in an attempt to reach as many searchers as possible. Unfortunately search engines do not provide the algorithms they use to rank content with many factors obfuscated to both users and marketers; this results in a lot of wasted effort when it comes to achieving those goals.

John Wanamaker famously said “Half the money I spend on advertising is wasted; the trouble is I don't know which half." This is a conservative estimate of the amount of effort wasted by search engine marketers when it comes to online advertising and promotion. I am aiming to use my capstone project to begin to identify actionable data to focus effort on the highest reward areas.

**PROJECT IDEA 1**

*What is the problem you want to solve?*

To identify which search engine ranking factors are relevant to ranking success.

*Who is your client and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn’t have otherwise?*

Client would be agencies who work with Search Engine Optimisation (SEO) and identifying where to spend time improving sites results in less wasted effort.

*What data are you going to use for this? How will you acquire this data?*

Proprietary data sets collected by in house technology and utilising 3rd party data from providers such as ahrefs, moz, majestic, pulsar and sharedcount. Historical data to help identify effect of changes may also be available.

Data is already collected but will need to be extracted and sorted or data expressly for this purpose can be obtained using proprietary tech.

*In brief, outline your approach to solving this problem (knowing that this might change later).*

Analyse XX search engine results pages (SERPs) and initially identify in isolation any factors that appear to be correlated to search engine rankings. Then attempt to identify which in conjunction appear to be correlated to search engine rankings.

*What are your deliverables? Typically, this would include code, along with a paper and/or a slide deck.*

A process for identifying which search engine factors are most closely correlated to ranking.  
The code used to do so.  
A presentation on the process and findings.

**PROJECT IDEA 2**

*What is the problem you want to solve?*

To identify which platforms (URLs) rank for certain types of queries.

*Who is your client and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn’t have otherwise?*

Client would be agencies who work with Search Engine Optimisation (SEO) and identifying where to spend time creating sites results in less wasted effort.

*What data are you going to use for this? How will you acquire this data?*

Proprietary data sets collected by in house technology and utilising 3rd party data from providers such as ahrefs, moz, majestic, pulsar and sharedcount. Historical data to help identify effect of changes may also be available.

Data is already collected but will need to be extracted and sorted or data expressly for this purpose can be obtained using proprietary tech.

*In brief, outline your approach to solving this problem (knowing that this might change later).*

Analyse XXX search engine results pages (SERPs) and initially identify in isolation which sites rank most prominently for each search type.

SERPs will be filtered in to search 'templates' (navigational, transactional, informational) and compared against other SERPs in of the same search type.

*What are your deliverables? Typically, this would include code, along with a paper and/or a slide deck.*

A process for identifying which platforms are most likely to rank for a given keyword.  
The code used to do so.  
A presentation on the process and findings.

**PROJECT IDEA 3**

*What is the problem you want to solve?*

Identify which On Page factors occur most commonly on the top ranking sites for any keyword.

*Who is your client and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn’t have otherwise?*

Client would be agencies who work with Search Engine Optimisation (SEO) and identifying where to spend time changing content results in less wasted effort.

*What data are you going to use for this? How will you acquire this data?*

Proprietary data sets collected by in house technology and utilising 3rd party data from providers such as alchemy.

Data can be obtained using proprietary tech and APIs.

*In brief, outline your approach to solving this problem (knowing that this might change later).*

Analyse XXX search engine results pages (SERPs) and initially identify in isolation which On Page factors each site has that are significant .

Results will have to be shielded against influence of factors not being tested and compared 'like-for-like'.

*What are your deliverables? Typically, this would include code, along with a paper and/or a slide deck.*

A process for identifying which On Page factors are most likely to lead to ranking.  
The code used to do so.  
A presentation on the process and findings.