Experiment Report – <name of experiment>

#### Experiment Owner:

#### Team:

#### Date:

Table of Contents

[Problem Description 2](#_Toc51081933)

[Proposed Solution 2](#_Toc51081934)

[Hypothesis 2](#_Toc51081935)

[Metrics 2](#_Toc51081936)

[Who or what is impacted 2](#_Toc51081937)

[Exceptions 2](#_Toc51081938)

[Experiment Configuration 2](#_Toc51081939)

[Results 2](#_Toc51081940)

[Human Evaluation Metrics 2](#_Toc51081941)

[All Queries 2](#_Toc51081942)

[Handled Queries 2](#_Toc51081943)

[Behavioral Metrics 3](#_Toc51081944)

[All Sessions 3](#_Toc51081945)

[Search Sessions 3](#_Toc51081946)

[All Queries 3](#_Toc51081947)

[Handled Queries 3](#_Toc51081948)

[Performance and Reliability Metrics 3](#_Toc51081949)

[Hypotheses result 4](#_Toc51081950)

[Conclusion & Insights 4](#_Toc51081951)

[Decision 4](#_Toc51081952)

# Problem Description

<what problem are you trying to solve?>

# Proposed Solution

<how do you propose solving it?>

<what are the tools/techniques involved?>

# Hypothesis

< An explanation of WHY you are suggesting this test. It could be from research, customer feedback, observed behaviours, analytics ...

Filling in these blanks is a good way to present a hypothesis and expectations of success:

Because we observed [data]and [feedback],

we believe that [changing something] for [a group of people] will result in [changed behaviour/outcome].

We’ll know this is successful when we observe [data] and [feedback]

>

# Metrics

<what are the metrics to measure to determine success>

# Who or what is impacted

<who (users groups) or what (queries) are impacted>

# Exceptions

<who are we excluding from the test or analysis and why?>

# Experiment Configuration

<what is the configuration of the test, how much traffic, when did you ramp, etc>

# Results

## Human Evaluation Metrics

### All Queries

#### DCG

### Handled Queries

#### DCG

## Behavioral Metrics

### All Sessions

<session metrics here>

### Search Sessions

<session metrics for sessions that have a search>

### All Queries

<query level metrics for searches, such as …>

#### DCG

### Handled Queries

<query level metrics for searches that were known to be different between a and b>

#### DCG

## Performance and Reliability Metrics

<metrics of speeeeeeed>

#### Performance

#### Reliability

## Hypotheses result

|  |  |  |
| --- | --- | --- |
| Hypothesis | Result | Supported? |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

# Conclusion & Insights

<what were the takeaways, what did we learn, what can be ported to other products, features, or algorithms>

# Decision

<this doesn’t get filled out until the launch review meeting. It should be agreed upon by all key stakeholders>