STATISTICS SAVES THE DAY

An ANLY482 Analytics Practicum Project

By Team Kyu-BI

Image credits: Joomag

OVERVIEW

Project introduction

Dataset description

Phase 1: The Predictive Model

Phase 2: A Revised Approach

Conclusions and learning journey

PROJECT INTRODUCTION

1. About SingPath

2. The Initial Proposal: The Predictive Model

- Assist instructors when conducting courses
- Build an understanding of questions

DATABASE DESCRIPTION

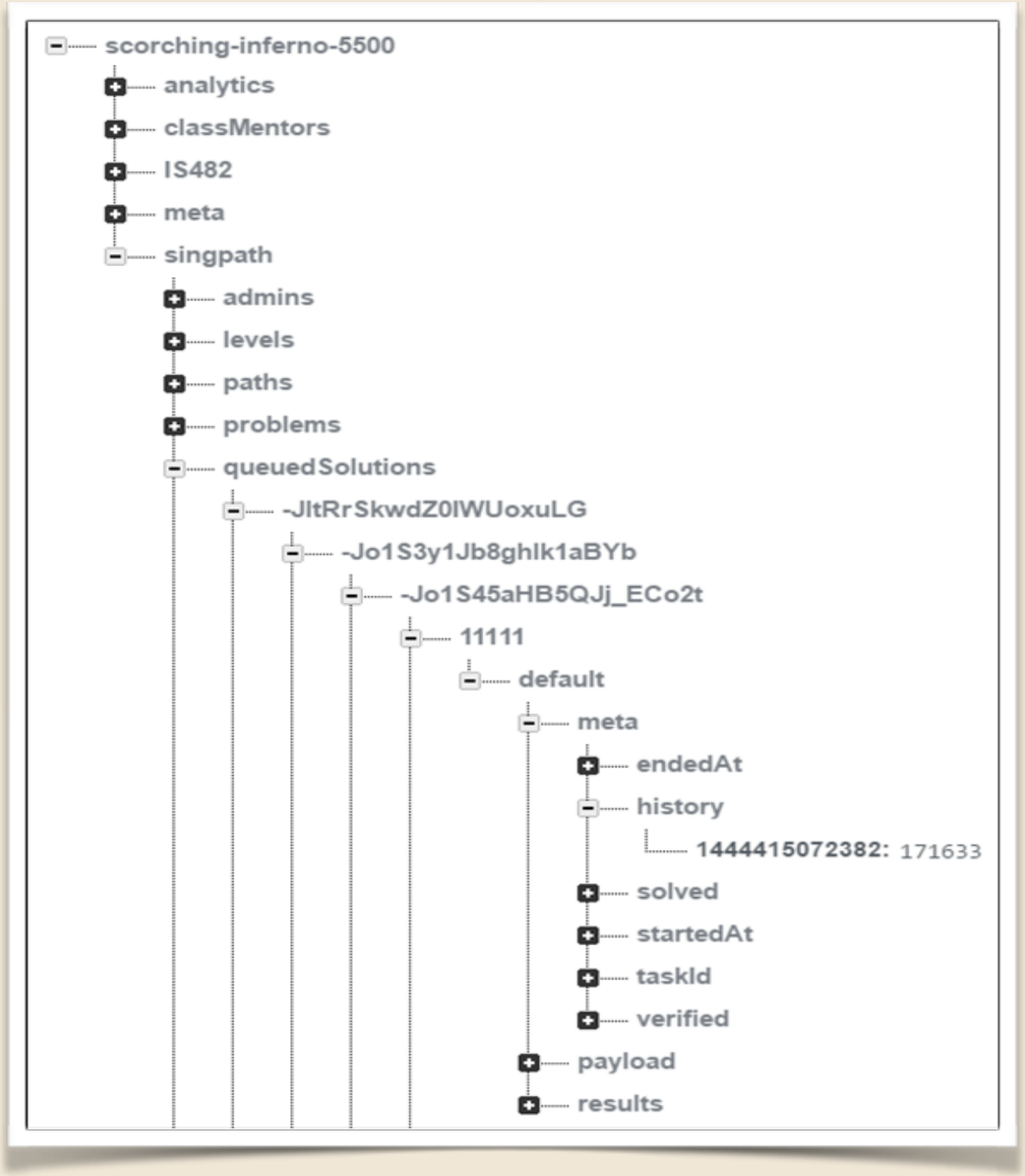
DATABASE

Firebase

BASIC INFORMATION

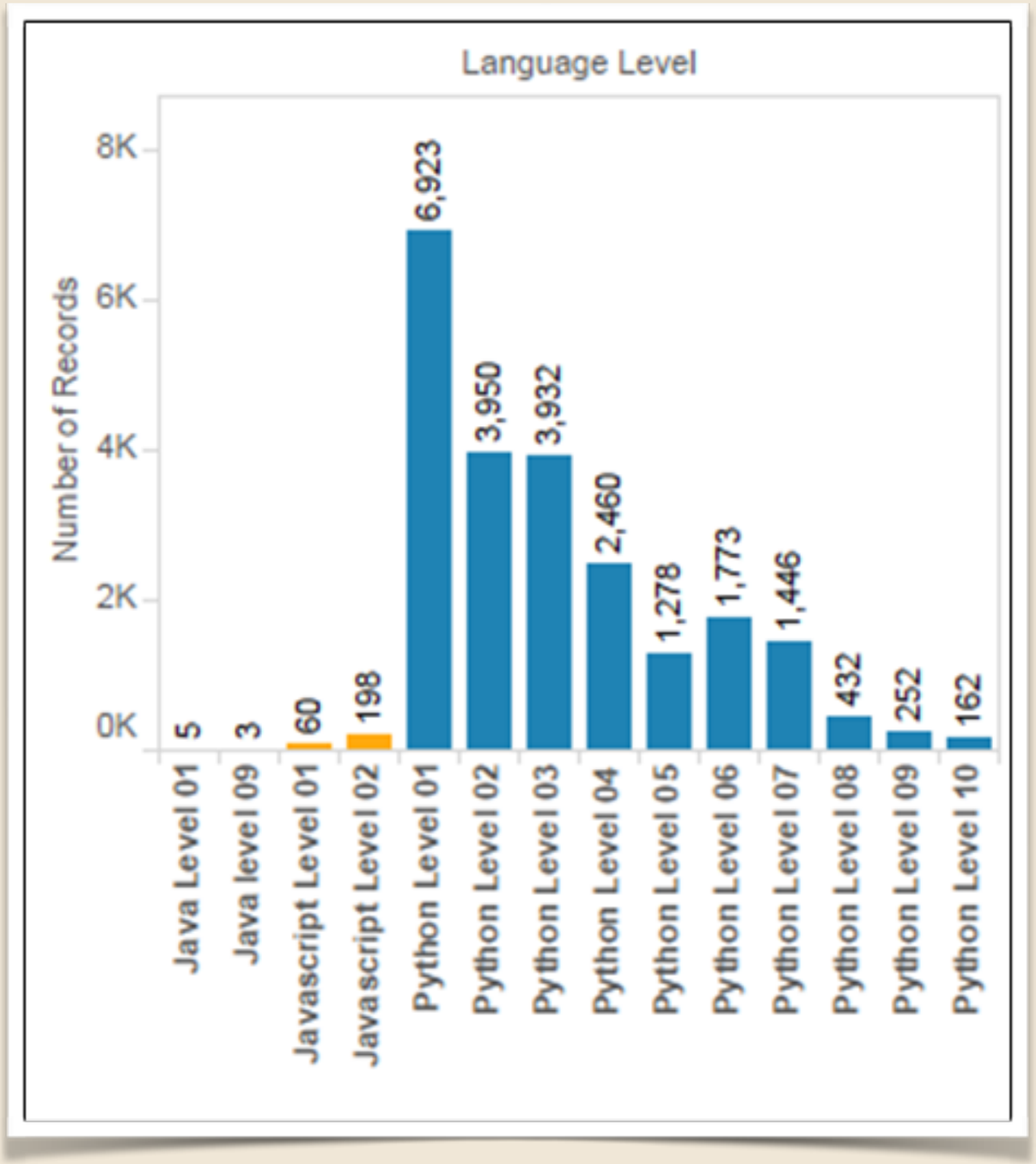
Unique records	22,874
Unique questions	223
Unique users	1,577

NATURE OF THE DATA



- Question Attempts
- Unique Language Key
- Unique Level Key
- Unique Question Key
- Unique User Key
- History of Attempts

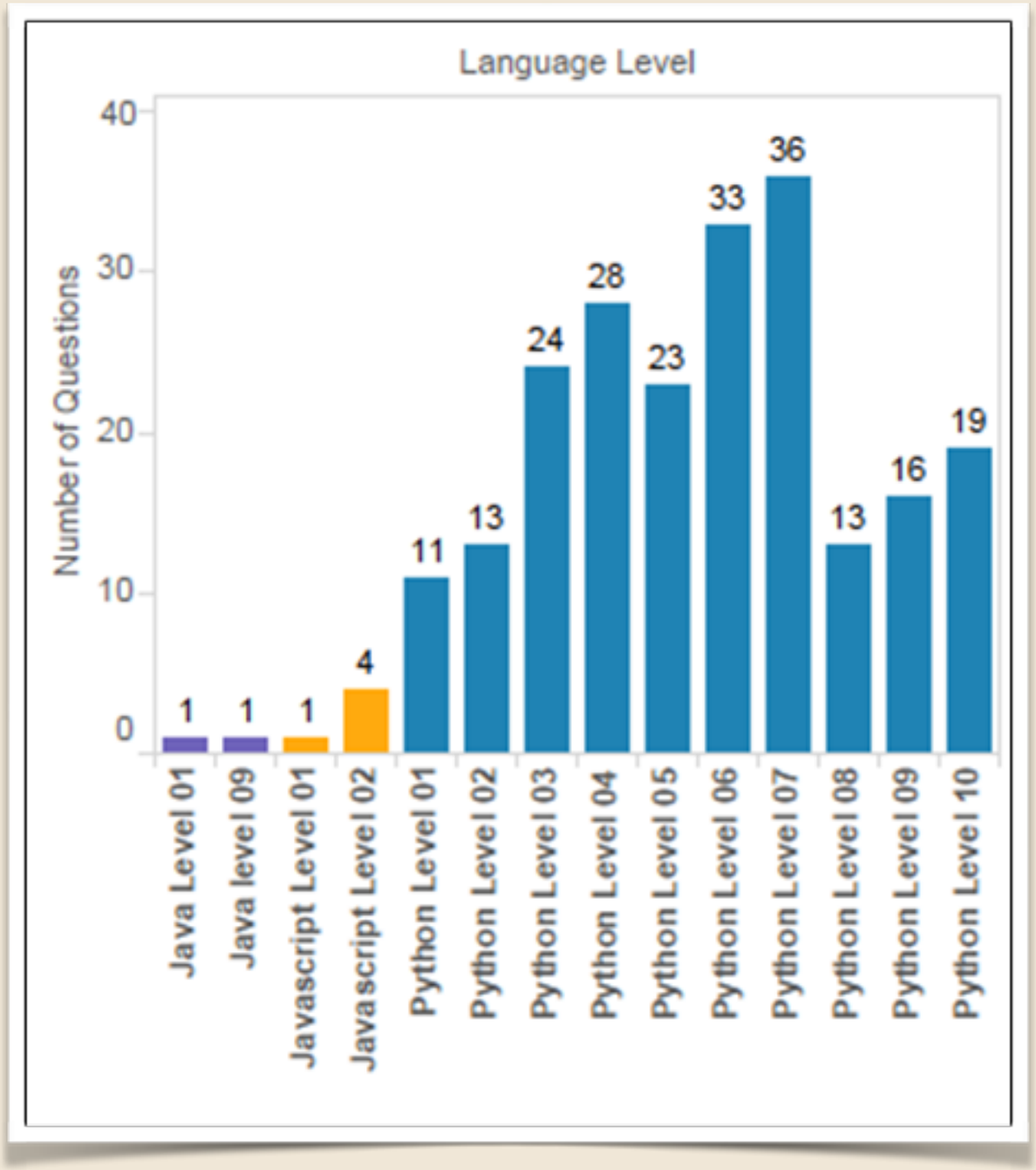
- ONLY 0.1% of attempts have any history!



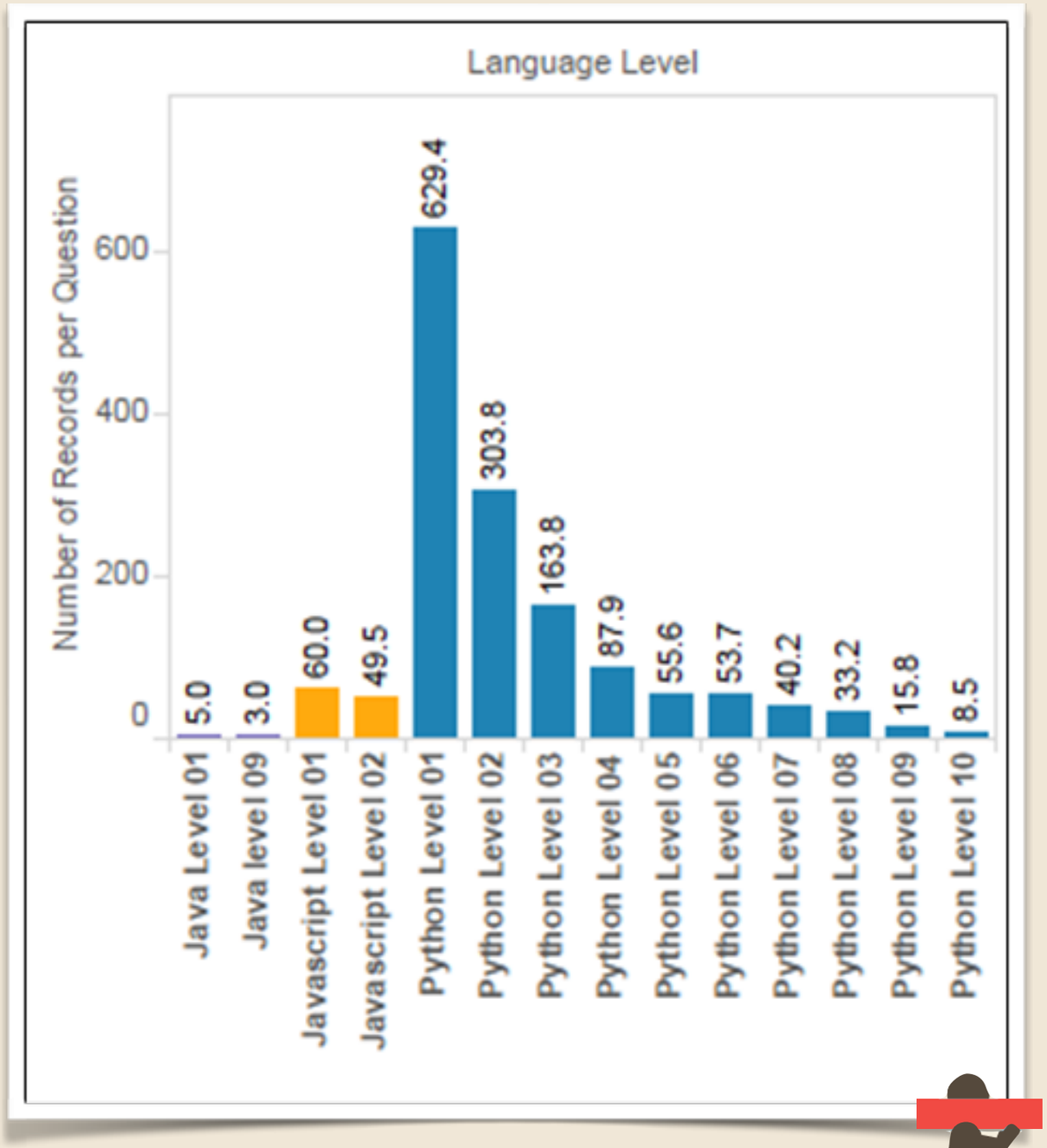
Number of attempts per language:

- Java = 8
- Javascript = 258
- Python = 22,608

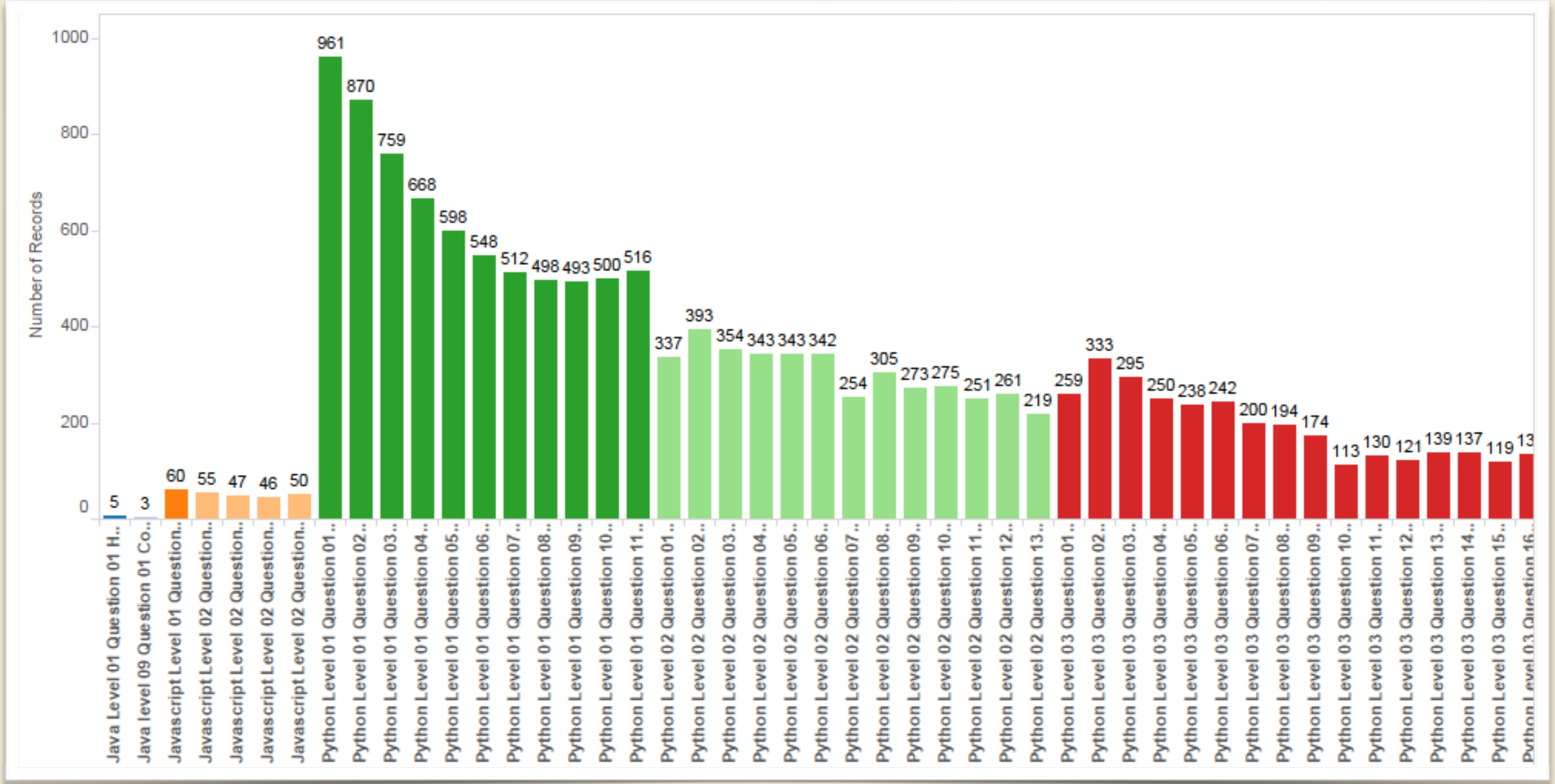
Number of Questions per Language Level



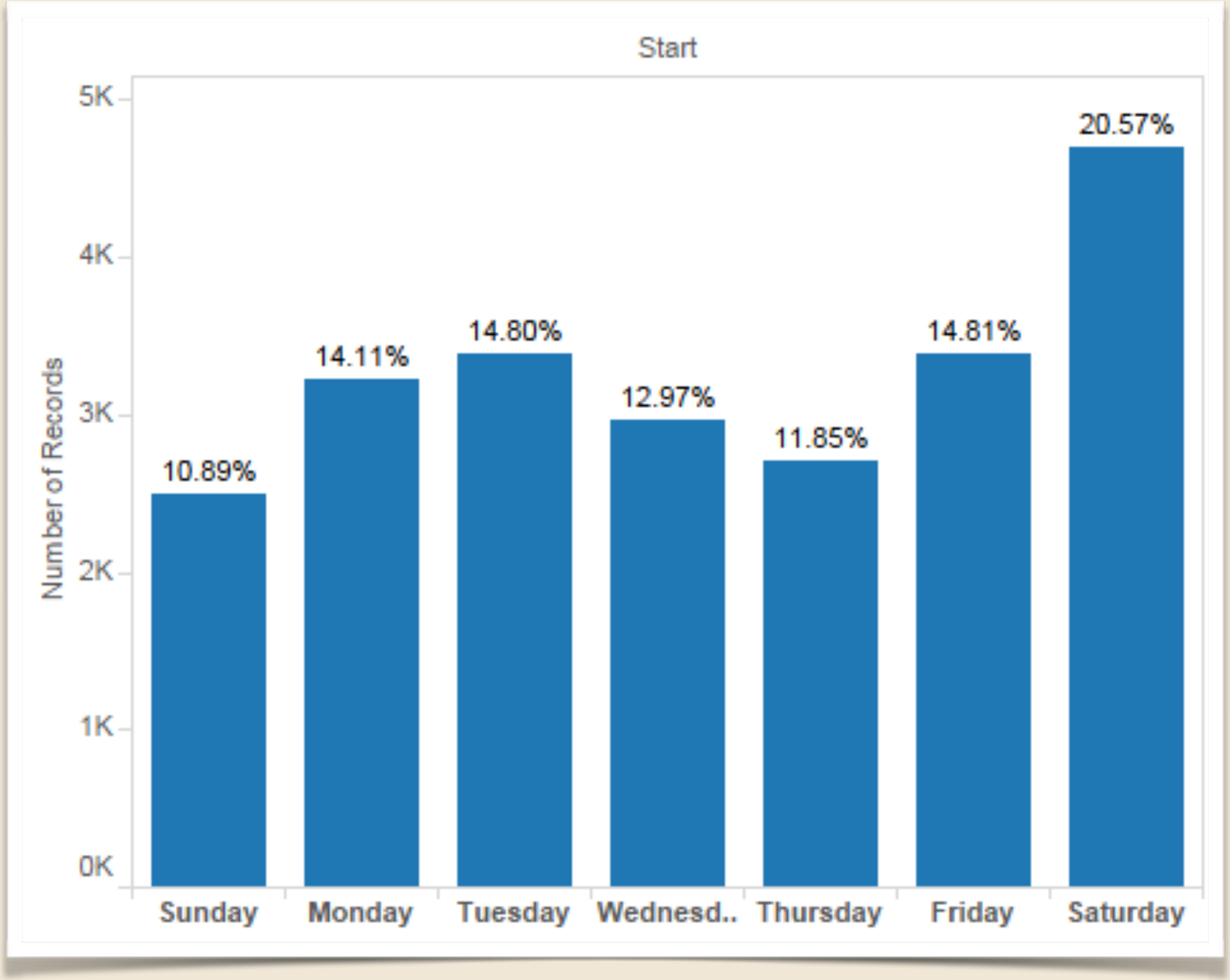
Average Attempts per Question per Language Level



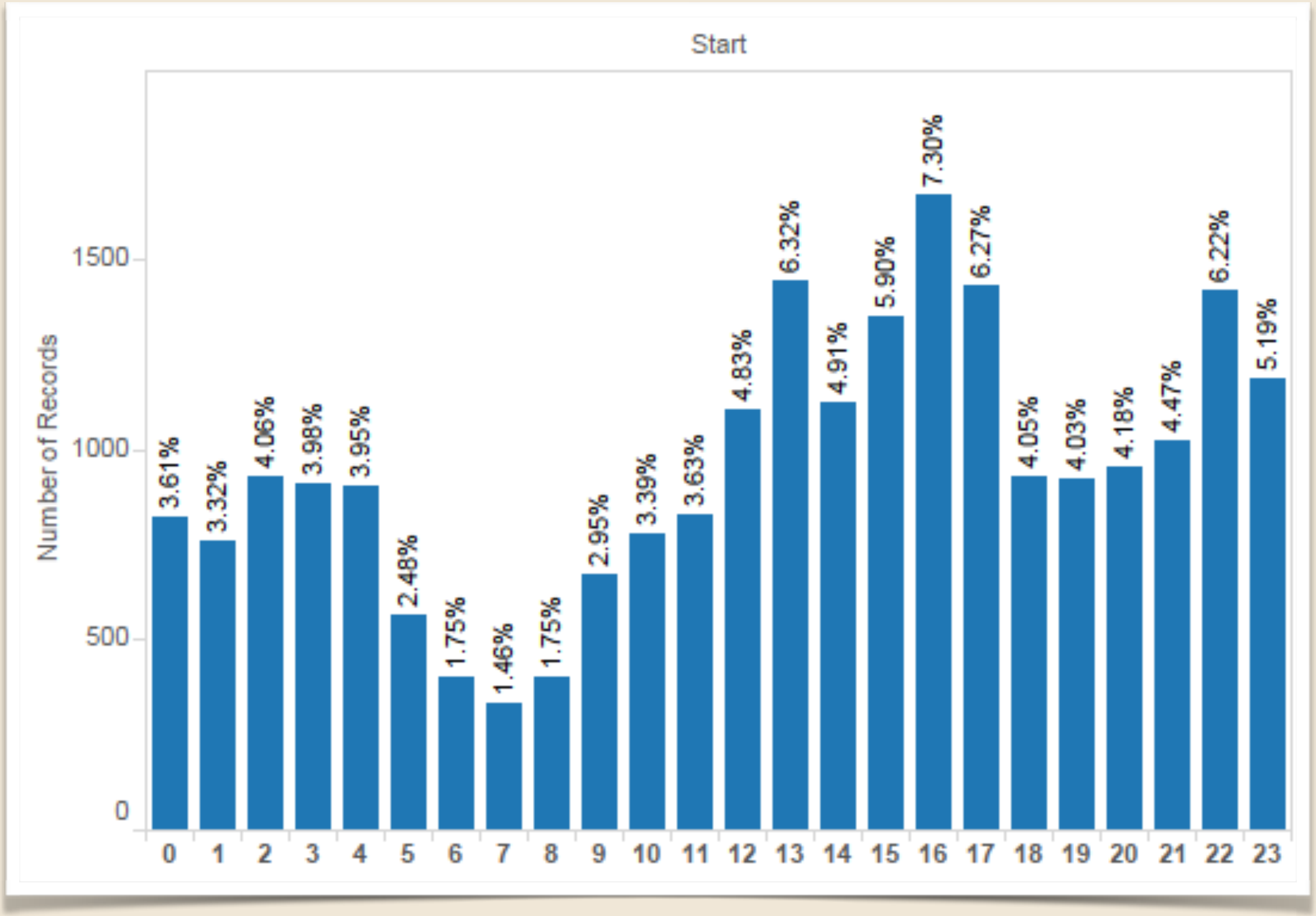
Question Attempts by Language Level



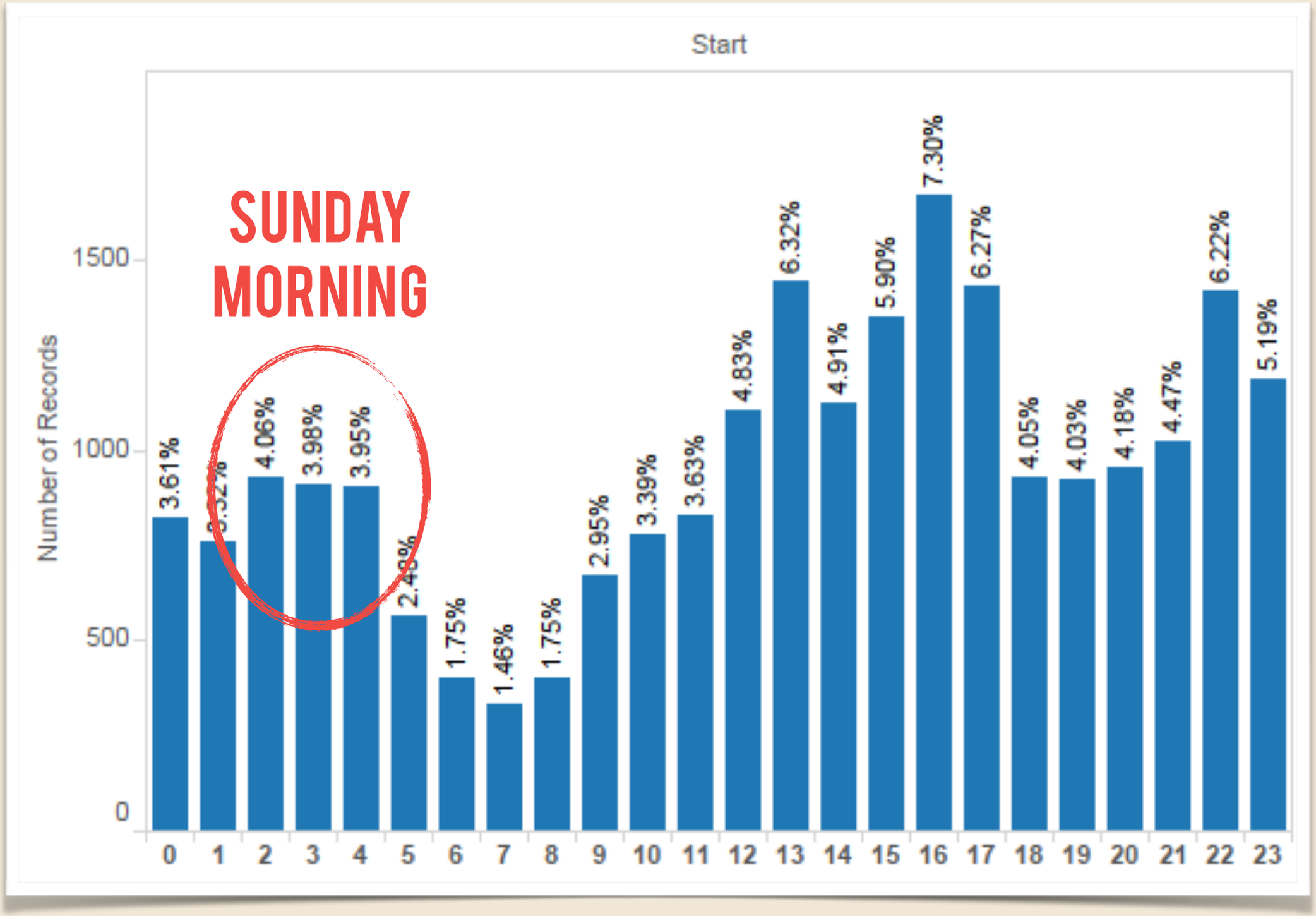
Attempts by Days of Week



Attempts by Hours of Day



Attempts by Hours of Day



THE PREDICTIVE MODEL

PHASE

01

THE IDEA

Use Means and Percentiles to predict completion time

WHAT WE WERE TRYING TO PREDICT

Time taken by a user to generate a correct answer using 2 methods:

(i) Median Time
(ii) Percentile Rank

MEDIAN MODEL

What is the Median time taken for all correct attempts on a question?

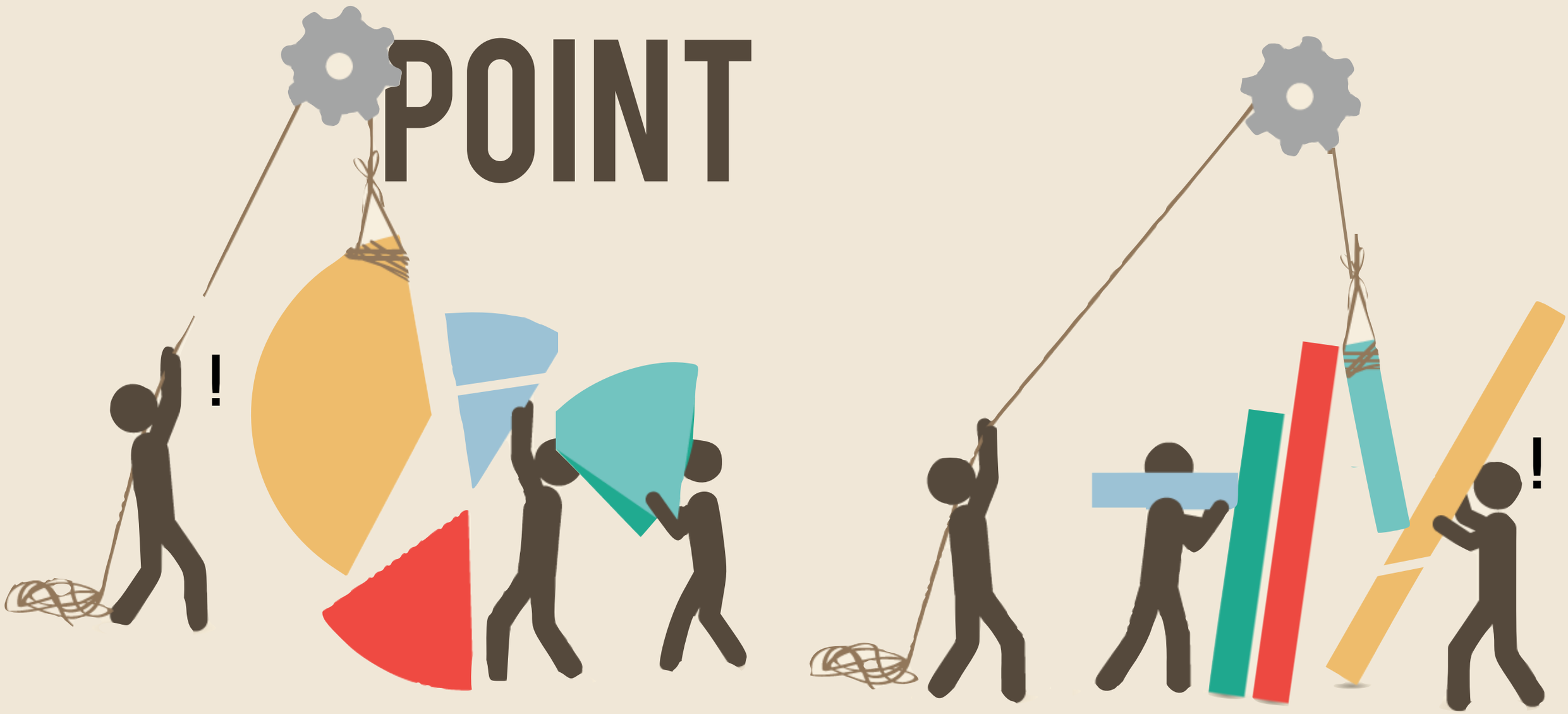
PERCENTILE MODEL

What is the average percentile this user falls into when attempting a question?

DEMONSTRATION OF MODELS VIA EXCEL

COMPARING THE MODELS WITH SSE

THE CRISIS POINT

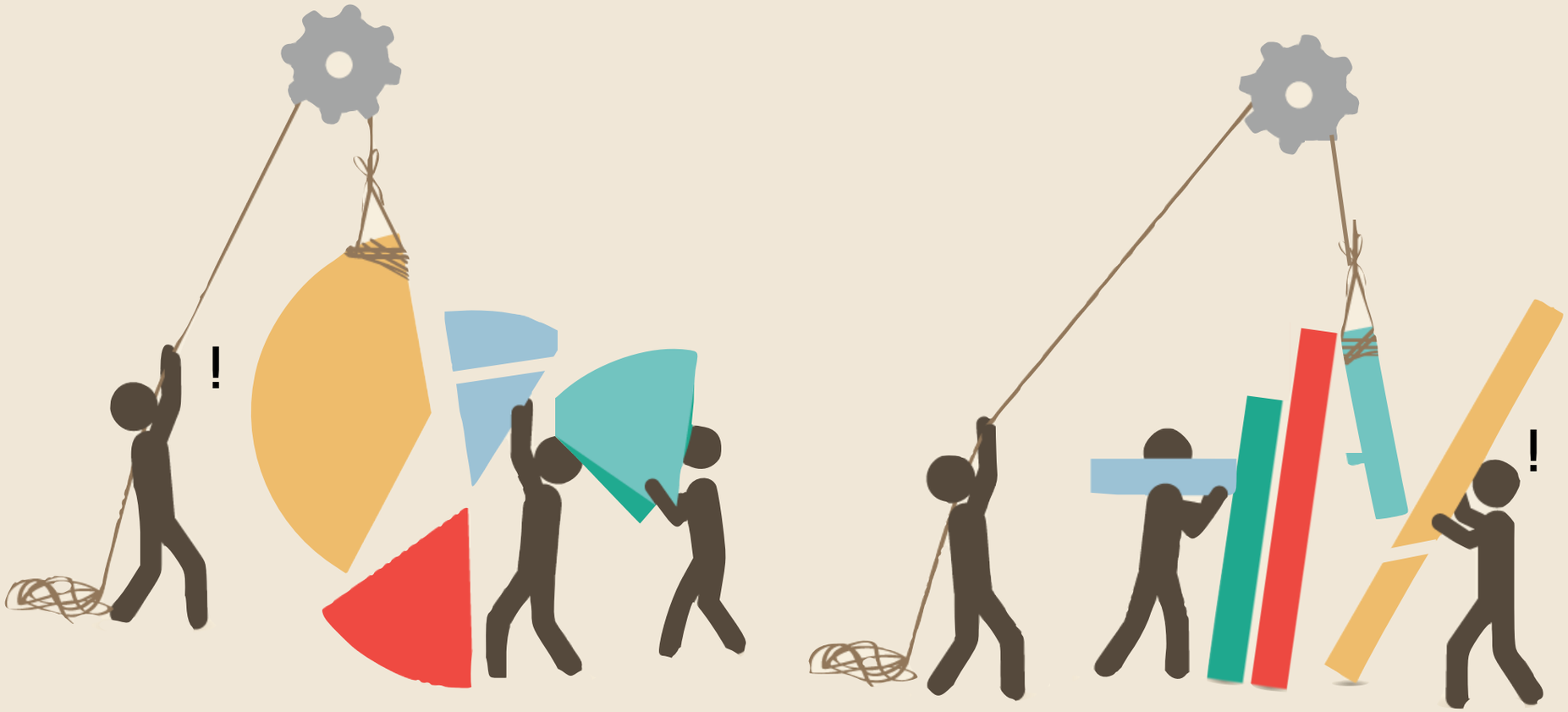


THE CRISIS POINT

Interim Review

Review of Predictive Models

New project needed

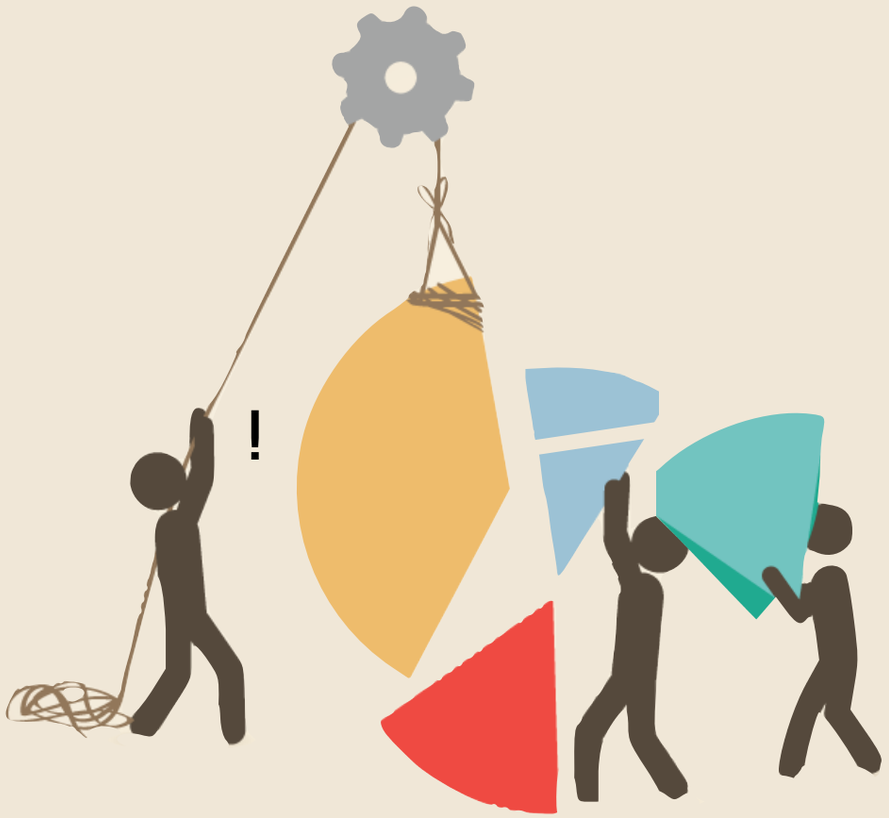


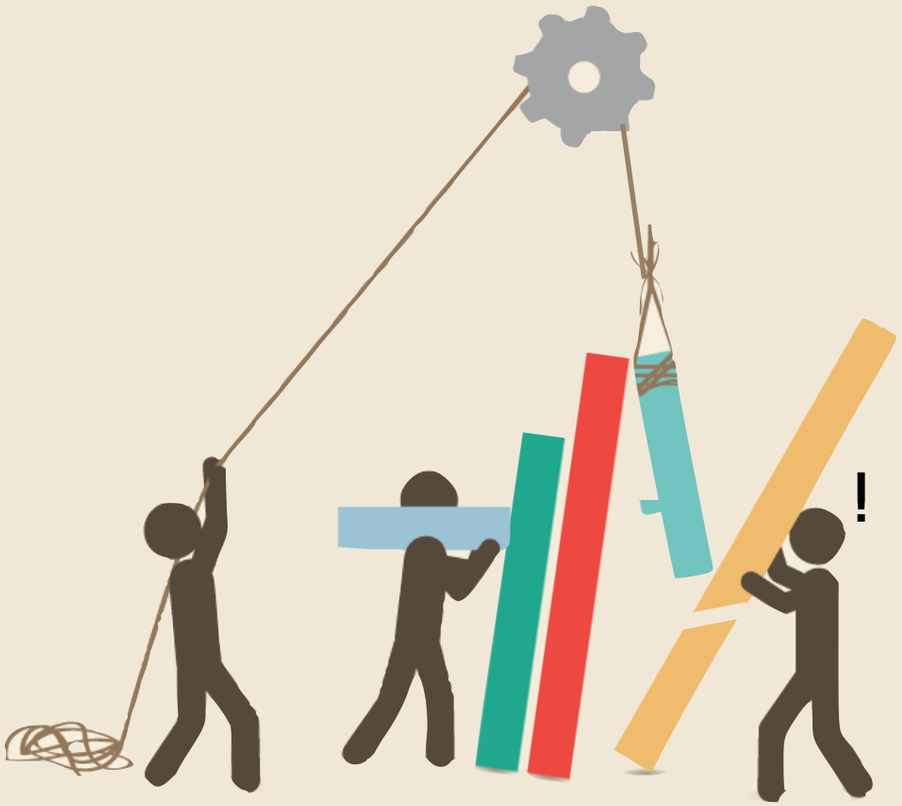
THE CRISIS POINT

All results < 10minutes

Data in rows were over-written

Very dirty data





A REVISED MODEL

PHASE

02

LITERATURE REVIEW

Self-directed learning

- Knowles, M. S. (1975). Self-directed learning.

Problem-based learning

- Smith, R. O. (2014). Beyond Passive Learning: Problem-Based Learning and Concept Maps to Promote Basic and Higher-Order Thinking in Basic Skills Instruction. Journal Of Research & Practice For Adult Literacy, Secondary & Basic Education, 3(2), 50-55.

LITERATURE REVIEW

- Self-directed learning
- Problem-based learning
- Experiment
  - Elgamal, A. F., Abas, H. A., & Baladoh, E. S. (2013). An interactive e-learning system for improving web programming skills. Education and Information Technologies, 18(1), 29-46.

1. REVIEW SINGPATH.COM

Attempted Python Levels 1 and 2

2. REVIEW DATASET

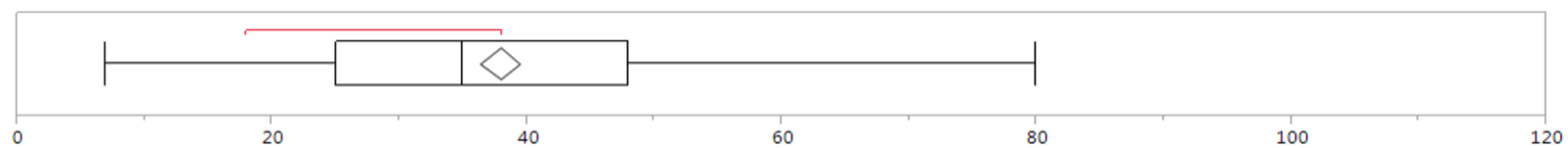
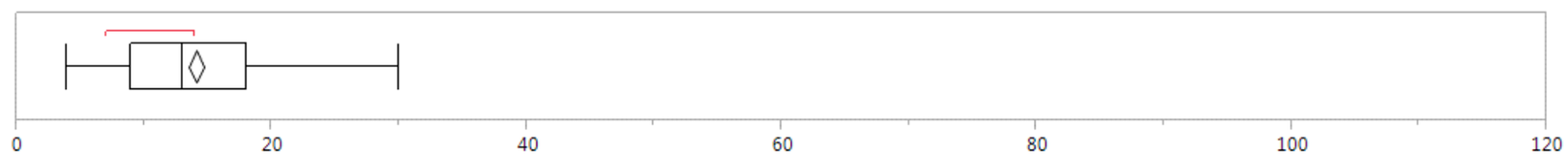
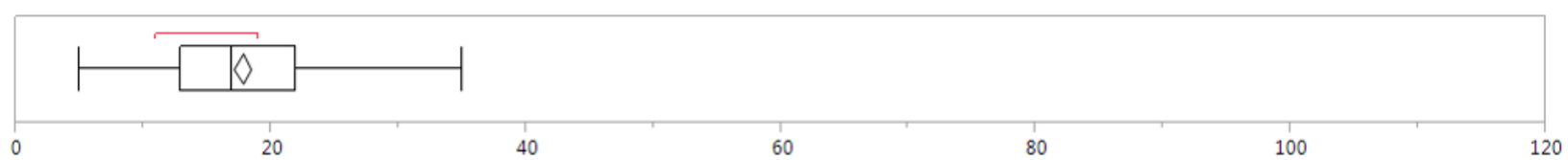
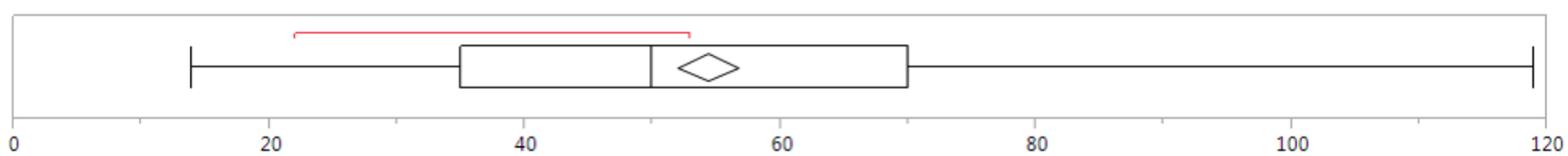
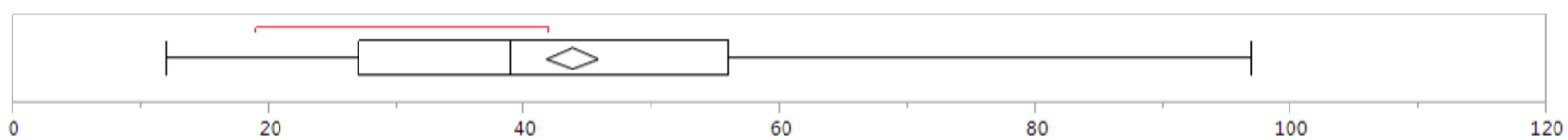
Revise data cleaning and analysis process

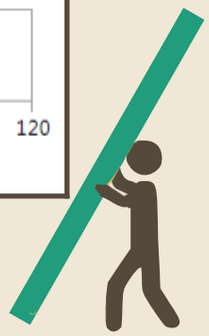
DEMONSTRATION OF SINGPATH

CURRENT SEQUENCE OF QUESTIONS

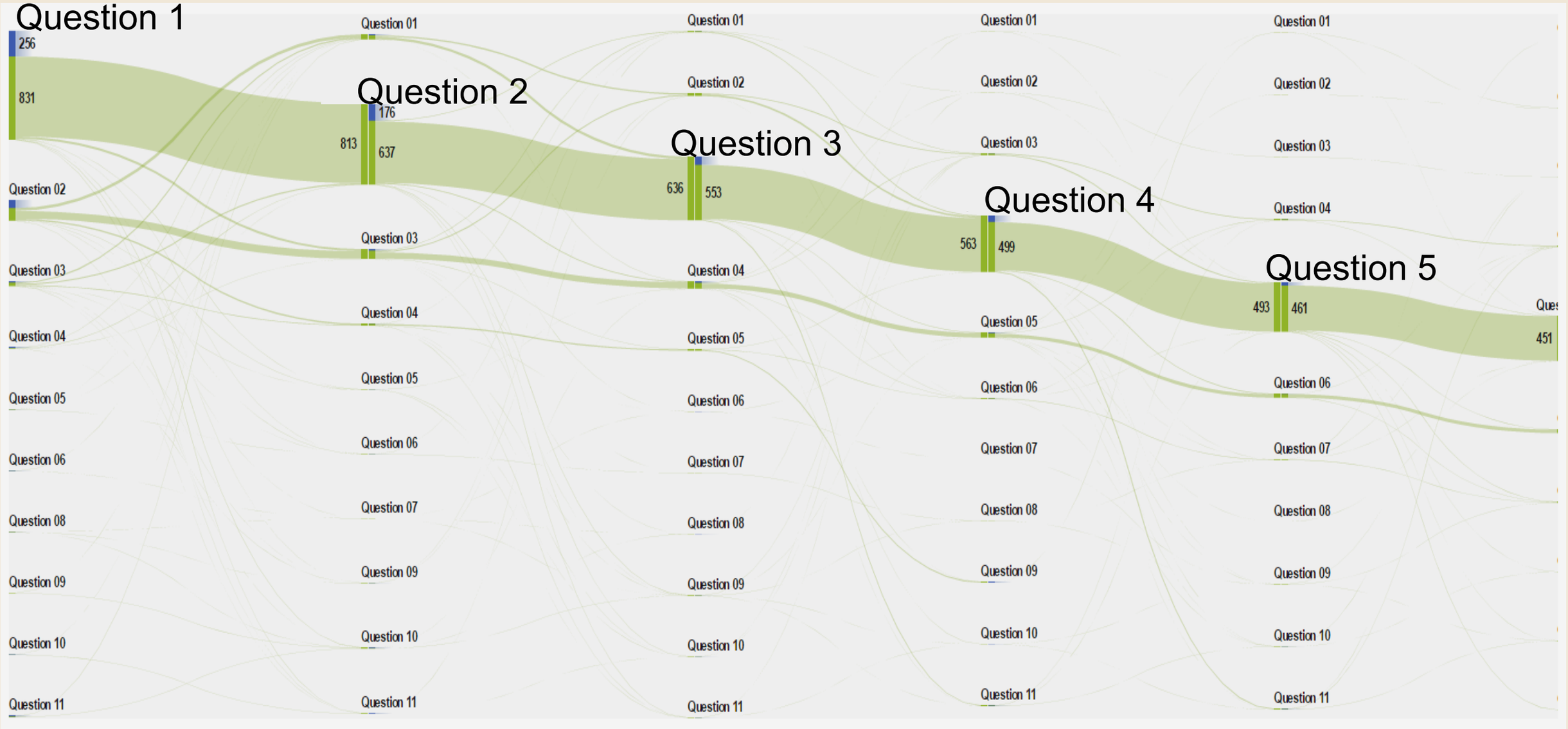
Question Number	Question Title
5	Still More Variables
6	Variables
7	Another Variable
8	More Fun With Variables
9	Many Variables

CURRENT SEQUENCE OF QUESTIONS

Question Number	Box-plot of Duration
5	
6	
7	
8	
9	



SELF-DIRECTED LEARNING METHOD

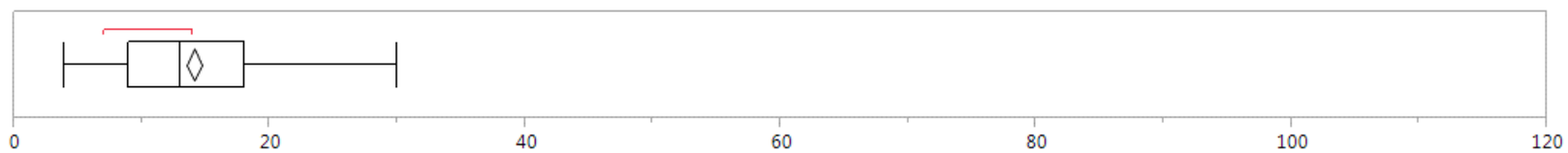
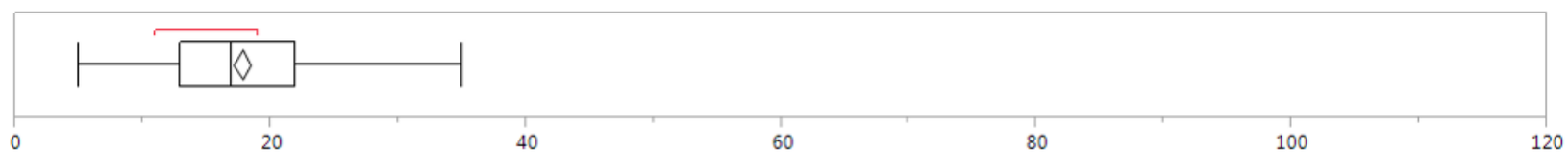
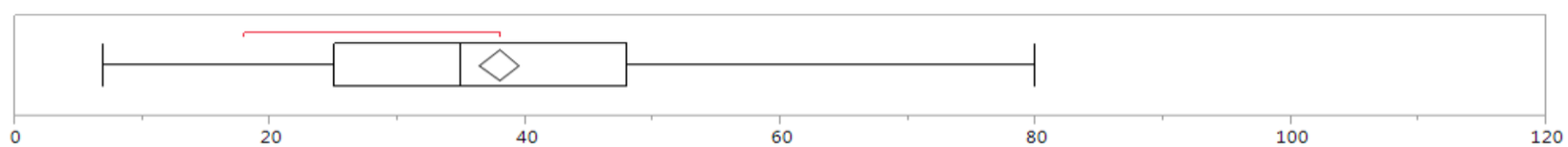
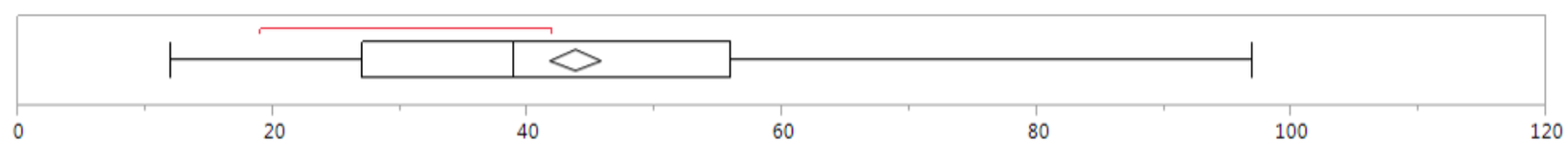
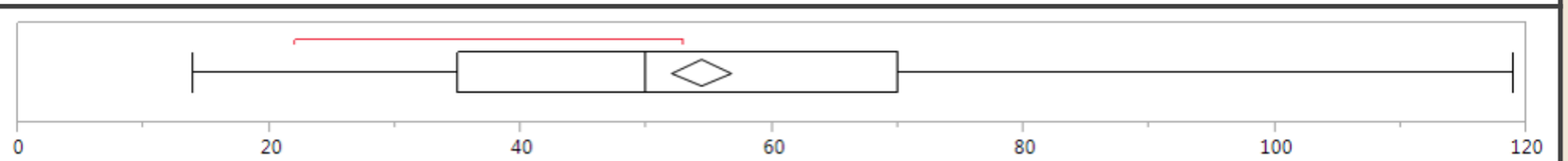


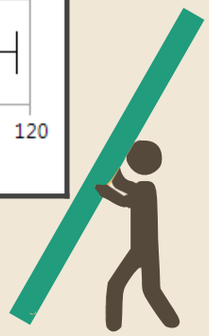
However, data indicates that users do not explore the questions in the order anticipated

PROPOSED SEQUENCE OF QUESTIONS

Question Number	Question Title
6	Variables
7	Another Variable
5	Still More Variables
9	Many Variables
8	More Fun With Variables

PROPOSED SEQUENCE OF QUESTIONS

Question Number	Box-plot of Duration
6	
7	
5	
9	
8	



LIMITATIONS

1. No fixed sequence of questions
2. No time limits
3. History of attempts not captured
4. No list of question competencies
5. Results of attempts not captured

RECOMMENDATIONS

1. Reorganise questions based on competencies learnt

2. Restructure database for future analytics

3. Update SingPath

EXPERIMENT

Compare different learning pedagogies

CONCLUSION: THE LEARNING JOURNEY

Began with an idea

Crisis due to overly ambitious goals

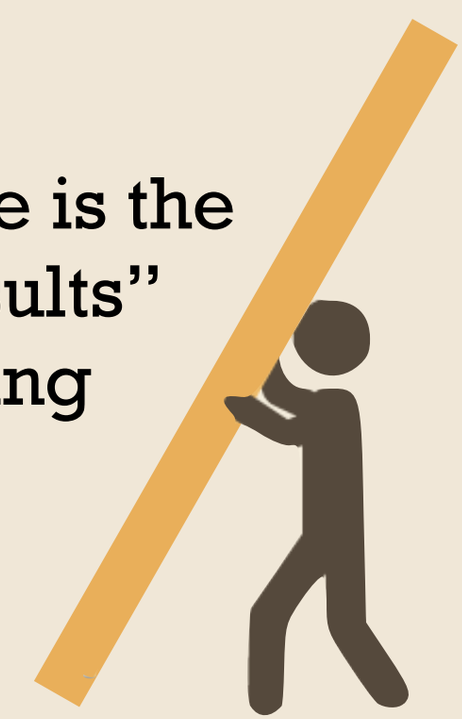
Simple analytical techniques achieved results

INDIVIDUAL REFLECTIONS

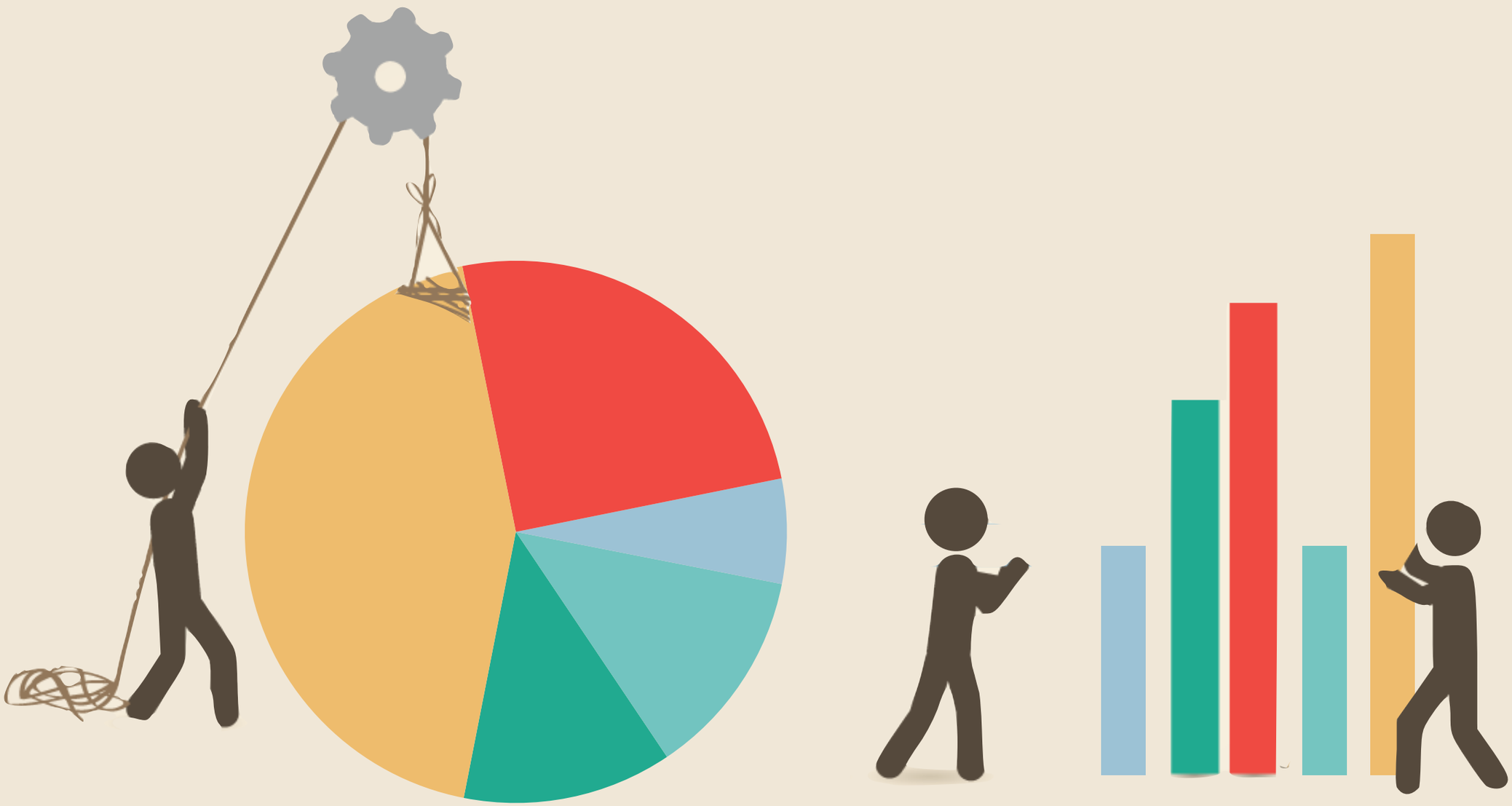
"Integrity and pride in a project helps move things forward" - Shane

"What does the data say?"
- Darren

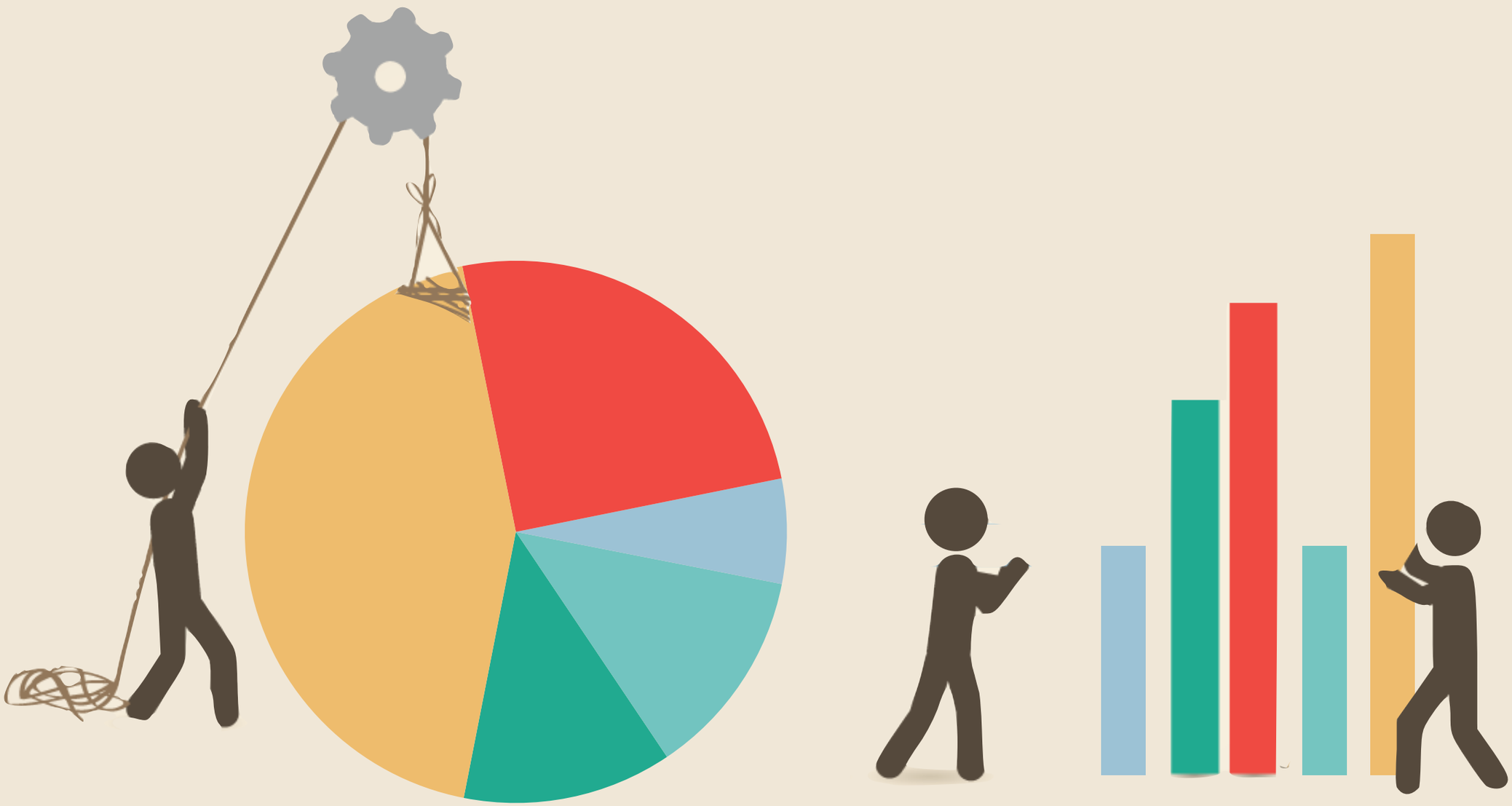
"Persistence is the key to results"
- Wei Yang



THANK YOU



APPENDIX



JSON TO CSV CONVERSION

- HTML file
- Pseudo-Code

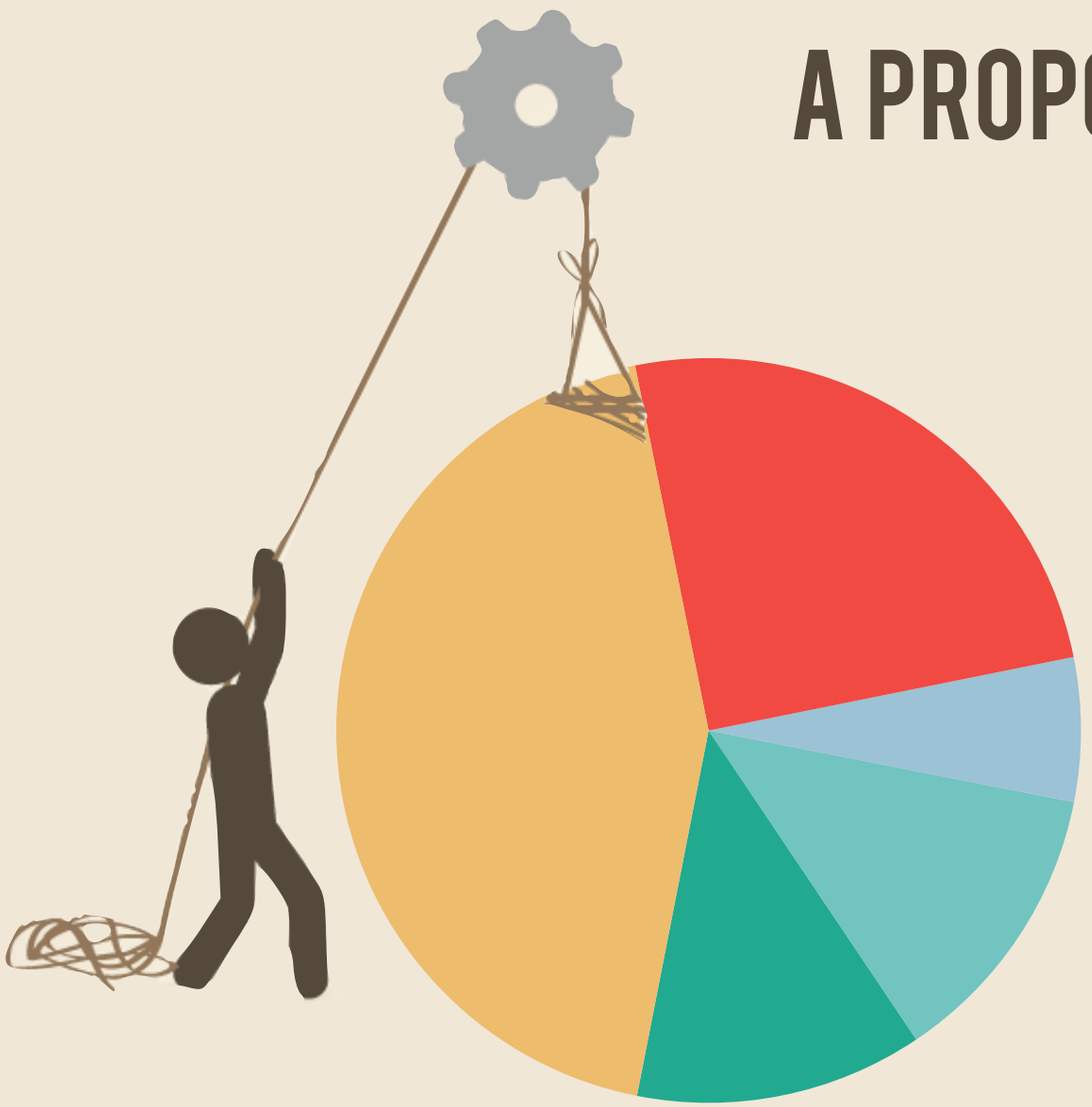
- For each language,
  - For each level of the language
    - For each user attempt of the question
      - Save the language level, question number, question title, username, start-time, end-time
      - Also look up and save the full titles of each question, including the level and language
      - Convert Unix time into a human-readable datetime format
      - Convert Duration attempts from milliseconds to seconds (rounded up to nearest second)

OTHER IDENTIFIED GROUPS

Language	Level	Group	Question Numbers	Proposed Sequence	Description
Python	1	Variables	5, 6, 7, 8, 9	6, 7, 5, 9, 8	This set of questions aims to educate the user on the use of variables
Python	2	Integer Data Type	3, 5, 6, 7, 8	3, 5, 6, 7, 8, 4, 13, 2, 9, 10	Begins with an introduction to the Integer data type, followed by basic application examples.
Python	2	String Data Type	2, 9, 10		Begins with an introduction to the String data type, followed by basic application examples.
Python	2	Float Data Type	4, 13		Begins with an introduction to the Float data type, followed by basic application examples.

EXPERIMENT DESIGN

A PROPOSED SOLUTION



EXPERIMENT INTRODUCTION

- Hypothesis
  - Students who complete a set of questions do better using the “Problem Based” pedagogy as compared to the “Self-Directed Learning” pedagogy
- Language to be learned: Python
- Participants
  - 2 Groups of Tertiary Students
  - Pre and Post Experiment Tests
- Control and Experiment Groups

EXPERIMENT DETAILS

- Sample of difference in question order

Language	Level	Original Question Order	Modified Question Order
Python	1	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11	1, 2, 3, 4, 10, 11, 6, 7, 8, 9, 5
Python	2	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13	1, 11, 12, 3, 5, 6, 7, 8, 4, 13, 2, 9, 10

EXPERIMENT ANALYSIS

- In each group, for each question
  - Take the results of those who got the question correct

- Compare Performance by Duration of Attempt

- Conduct 2-Sample 1-Tailed Z-Test
  - Or a 2-Sample 1-Tailed T-Test

EXPERIMENT MISSING COMPONENT

- Questions for the Pre and Post Experiment Tests
- Questions for the Competencies Test