

»» SURVEY PROPOSAL ««

PROGRAMMING LANGUAGE PREFERENCES AMONG
DATA SCIENCE STUDENTS:
PYTHON, R, AND JAVA

FACULTY OF COMPUTER AND DATA SCIENCE

18 APRIL, 2023

TEAM 11

**SURVEY
METHODOLOGY**

Dr. Ebrahim Galal

SUPERVISOR

Eng. Passant Hamza





ABSTARCT



INTRODUCTION

Python, R and Java languages are the most popular programming languages used for various purposes. They have their own unique features and functionalities that make them suitable for different tasks.

This survey aims to determine the preferred programming language among computer and data science students. The survey will gather information on participants' experience with each language, as well as the factors that influence their choice.

Insights gained from the survey will inform educational institutions about the suitable programming language for their specifications and education level, leading to better educational outcomes and better-prepared graduates.





OBJECTIVES & HYPOTHESIS



OBJECTIVES

- Which language is preferred by the students?
- Which language is most used by students depending on their specification?
- What are the functionalities that affect their choice?
- How do students find the ease of learning and using each language?
- Rating students' familiarity level and experience with each language?
- How do job market demand for each language influence the language choice for students?
- How do students choose between these three languages for specific data science tasks or projects?
- What is the importance of having knowledge of multiple languages in the field of data science?

HYPOTHESIS

- The preferred language between the students is Python.
- There's a correlation between age and preferred language
- There's a correlation between level and preferred language
- There's a correlation between familiarity and preferred language
- There's a correlation between the gender and preferred language



PARTICIPANTS



PARTICIPANTS

The participants in this survey will be General department of Computer and Data Science faculty students.

The survey will be sent to 600 students of different levels and ages.

Recruitment:

Participants will be recruited through various channels, including email, online communities, social media, face-to-face interactions. With a Thanks card to encourage participation.

Survey Method:

The survey will be conducted using Google Forms to gather the required information with a combined open-ended and closed-ended question.



SURVEY DESIGN



SURVEY DESIGN

The survey will consist of closed-ended questions, including multiple choice, ranking questions, binary questions, and checklist questions with 'others' option so the student can fill it with different answer. It will also consist of open-ended questions to give students the freedom to express their opinions.

The survey questions will be designed to be clear and easy to understand, and will be pilot tested with a small group of students before being distributed to the larger sample.

The survey will be voluntary, and participants will be informed about the purpose of the survey and how their data will be used. The survey will take approximately 5-10 minutes to complete, and participants will have the option to skip any questions they do not wish to answer.

Overall, our survey design will be focused on gathering insights about preference of the three languages among these students.





DATA ANALYSIS

The Data analysis plan will consist of several steps to ensure that we draw accurate and relevant conclusions from the survey data.

Data cleaning and preprocessing:

The survey responses will be first cleaned and prepared to be ready for analysis.

1. *Data Cleaning:*

- checking for missing and inconsistent responses
- Remove outliers

2. *Data Transformation:*

- encode open-ended responses.
- convert the ranking question into a numerical format.

Analysis Techniques:

1. *Descriptive statistics:*

using measures of central tendency and variability to provide an overview of the data (e.g. frequency, mean, standard deviation, ...)

2. *Inferential statistics:*

Applying regression analysis, t-tests to make inferences about the population based on the sample data (e.g. hypothesis test)

3. *Text Mining:*

Apply text analysis techniques for open-ended responses (e.g. sentiment analysis, ...)



DATA ANALYSIS



Data Visualization:

Data visualization techniques are to represent the data derived from the analysis in a way that makes it easy to understand. This helps to identify patterns and trends in the data, which can be communicated to stakeholders in a clear and straightforward manner to draw better insights.

Some of this visualization techniques:

1. *Bar charts:*

To display categorical data by showing the frequency of responses for each category.

2. *Pie charts:*

To show the proportion of responses for each category.

3. *Scatter plots:*

To display the relationship between two numerical variables, and also using the size of bubbles to represent a categorical variable (bubble chart).

4. *Word clouds:*

To display the most common words or phrases in the open-ended survey responses.

Interpretation of Results:

compare the results obtained from the data analysis with the initial expectations. And identify the most important and significant findings from the analyzed data.





THE TIMELINE



TIMELINE

The survey project will be completed in four weeks, divided into seven phases:

APRIL 12
APRIL 17



Phase 1: Proposal

Outlining the importance and goals of the survey, specifying the main points of the whole process of making a survey, and determining why the survey is needed.

Phase 2: Questionnaire

The survey questions will be carefully crafted to ensure that they are relevant and clearly written. To collect information and data that will answer the research objectives of the survey.

APRIL 26
MAY 9



Phase 3: Sampling and Pretests

defining a sample group based on characteristics and conducting pre-tests to identify and solve problems with the survey before giving it to the larger group.

Phase 4: Collect the Data

collecting the survey responses and preprocessing the data to ensure completeness and accuracy and that it's ready for analysis.

Phase 5: Expectation vs Real

compare the expected outcomes to the actual results.

To evaluate the effectiveness and ensure that the survey project meets its objectives and provides valuable insights

MAY 10
MAY 20



Phase 6: Data Analysis

Analyze the survey data to interpret the survey data and provide insights. using different statistical techniques and tools.

Phase 7: Report & Presentation

Communicating the findings of the survey based on the results to be informative and accessible to the audiences.



THE BUDGET



BUDGET

- **Software and tools:**

The survey will be conducted using Google Forms, which is a free online tool. There may be additional paid software or tools required for preparing and designing documents or in data analysis.

- **Incentives for participants:**

To encourage participation, a small incentive such as a gift card or a prize can be provided.

- **Miscellaneous expenses:**

For any other requirements such as printing or phone calls.



ETHICS AND CONFIDENTIALITY



ETHICS AND CONFIDENTIALITY

Adhering to ethical guidelines in the conduct of this survey to ensure participant confidentiality.

By implementing the following measures:

- **Informed consent:**
Participants will be provided with an explanation of the study's purpose and procedures, and will be given the opportunity to provide informed consent before participating in the survey.
- **Anonymity:**
Participants will be assured that their responses will be anonymous. personal information will not be shared with anyone outside of the research members.
No identifying information will be collected.
- **Confidentiality:**
All data collected during the survey will be stored securely and accessed only by research members.
- **Data protection:**
We will comply with all relevant data protection regulations and guidelines, and ensure that the data is stored and transferred securely.



CONCLUSION



CONCLUSION

In summary, the proposed survey to explore data science students' experience with Python, R, and Java programming languages is a well-designed approach to gain valuable insights.

The survey can help students make informed decisions about which language to learn based on their interests and specific data science tasks or projects.

And also it is critical in promoting better educational outcomes and preparing students for the job market in data science. As it can provide insights to educational institutions to improve their curricula.

Ultimately, this survey can help improve the field of data science by giving insights into what students need and what their experiences are.

This will better equip students with the necessary skills and knowledge to succeed in their future careers.

