**Sampling Plan for Research to Determine if Previous Computer-related Qualifications/Experience Affect Students’ Data Science Grades at Moringa School**

**1. Sampling Strategy**

**Objectives and Reliability Requirements**

* To determine whether a background in Computer science-related fields influences the data science performance at Moringa.
* The sampling data will cover a period in which cohorts 1 -13 studied at Moringa School.

**Target Population**

* The target population is the alumni students of Moringa School who studied the Data Science course.
* The target group comprises students who got admissions to Data Science courses at any level i.e core and prep.

**Sampling Approach/Method**

* Systematic Sampling: For every cohort, will do a simple random sampling to ensure all cohorts are represented. Also, will have equality in terms of gender (ratio 1:1).
* Cluster Sampling: The dataset will be divided into two clusters i.e Prep and Core

**Sample Size**

* Will have 10 males and 10 females per cohort. The total per cohort will be 20, giving us a total of 260 sample population.

**Sample Representativeness/Sampling Frame**

* The population of interest is all alumni Data Science students at Moringa School; the sampling frame includes only those students who satisfy all the following conditions:
  + The core grade and prep grade can be found from Moringa School exams department
  + The gender is listed in the Moringa School Admissions directory
  + The student can either be found through LinkedIn or Phone call to retrieve the past academic qualifications information.
* The population data will be kept in an excel spreadsheet and each row assigned a randomized number.
* The random numbers will then be selected using the systematic approach where for every cohort, 10 females and 10 males will be selected to form the sample population.

**2. Data**

**Data Collection and Field Measurements**

* The dataset will have the following columns:
  + Student Name - *object*
  + Student Gender - *numeric*
  + Prep Grade - numeric
  + Core Grade - *numeric*
  + Computer Science Background? - *numeric*
* Moringa school Admissions and Exams departments will be requested to provide details for Student Name, Student Gender, Prep Grade, and Core Grade.
* Procedure for data collection:
  + Request Student Name, Student Gender, Prep Grade, and Core Grade data from Moringa School.
  + Merge and clean data to eliminate missing values and anomalies.
  + Assign random numbers and select a sample.
  + Search for the students’' LinkedIn addresses
  + Search under “Education” and “Experience” and verify if the student had prior computer science or IT-related academic or work experience prior to enrolling to Moringa School for Data Science.
  + If the student has such experience, in the “Computer Science Background?” column in the sample dataset, assign 1, if not assign 0.
  + For the students with no LinkedIn addresses or such information is missing, request phone contacts from the Admissions Department and reach out to them and request for the information.
  + If the student has such experience, in the “Computer Science Background?” column in the sample dataset, assign 1, if not assign 0.
  + For the students whose data cannot be gotten through Linkedin or Phone calls, leave it as null.

**Quality Assurance / Quality Control**

* To ensure high quality of data collection, the following criteria will be followed:
  + Ensure all gender details are recorded and transformed to numeric. Males to 1 and female to 0.
  + Ensure all the previous qualifications are recorded and if any is computer science-related, record as 1 and if not, 0.
  + Ensure that the Prep grade and Core grade are recorded
  + Drop all rows with at least 1 null value

**Analysis**

* Find the correlation between Prep grade and gender
* Find the correlation between Core grade and gender
* Find the correlation between Prep grade and Core grade
* Find the correlation between Computer Science Background? and Prep grade
* Find the correlation between Computer Science Background? and Core grade
* Find the correlation between Computer Science Background? and average grade (average of prep grade and core grade)

**3. Implementation**

**Plan**

* This plan will be implemented by the Admissions department. The Lead Data Scientist at the department will lead the data collection and analyses.

| Weeks | Process | Remarks |
| --- | --- | --- |
| 1-2 | Data collection, retrieval, and cleaning |  |
| 3 | Data analysis |  |
| 4 | Data visualization and presentation |  |