Step 1

Human cognitive trait: Aspiration for success

Aspiration in general means a person's ambition to achieve success. According to Valeriu (2019), aspiration involves from the cognitive point of view the representation for achieving more (Valeriu, 2015). Further, self-efficacy and motivation are closely associated with aspirations (Bokan et al., 2020). Aspirations are used in the literature to predict success. For example, teenage career aspirations have been used to predict future earnings (Ashby & Schoon, 2010). The goal of this study is to measure data scientists' aspirations in the context of data science contests' communities like those on Kaggle (*Kaggle: Your Machine Learning and Data Science Community*, n.d.). To give a little bit of context on Kaggle, Kaggle is the world's largest data science community where data science competitions are held every day. After participating in a data science contest held on Kaggle, we will hand out a survey where we will ask data scientists to self-report on the relevant actions that will measure the aspirations of the data scientists for success.

Step 2

The main goal of my study is to measure the level of aspiration for success among data scientists. Throughout the literature, personality traits similar to aspirations have been widely studied, which includes motivation and ambition to examine their effects on organizations and individuals.

A similar trait that has been studied is ambition. Ambition is defined as "the persistent and generalized striving for success, attainment, and accomplishment" (T.A. Judge, J.D. Kammeyer-Mueller, 2012). In the article titled "Ambitious employees: Why and when ambition relates to performance and organizational commitment" published by Hirschi et. al. in 2021, the authors collected survey responses from 157 respondents on self-rated ambition and proactivity (Hirschi & Spurk, 2021). The authors also sent out surveys to the respondents' partners or supervisors to measure promotability and job performance (Hirschi & Spurk, 2021). The goal of this study is mainly to identify ambitious individuals and how they affect the organization they work in. A core finding of this paper is that ambitious individual is often on the lookout for interesting opportunities are more likely to resign voluntarily.

Another similar trait that has been studied is motivation in workplaces. An article titled "The Measurement of Employee Motivation by using Multi-Factor Statistical Analysis" written by Zamecnik in 2014 used a survey that adopted the 5-Point Likert Scale to obtain data about the motivation of employees in enterprises. The employees were asked to rate the 27 motivational factors given to them. In this study, the authors find that they might be possibilities whereby they could create unique motivational programs that cater to each unique group of employees based on the motivational values that they cherish using cluster analysis (Zámečník, 2014). This piece of research somehow relates to my study in the sense that the authors are attempting to find patterns from survey responses to see what are the thinking patterns that contribute to individual success.

The third paper that I have examined that relates tightly to aspirations titled "Do high aspirations lead to better outcomes? Evidence from a Longitudinal Survey of Adolescents in Peru" used educational, occupational, and aspirations to migrate to understand how teenagers persist towards success over time. The authors collected survey results from 400 teenagers in Lima, Peru (Graham & Ruiz Pozuelo, 2021). The important finding of this study is that higher aspiration level leads to positive behavior such as more efficient time use and better overall well-being of the students.

Step 3(a)

The population intended to measure is undergraduate and graduate students majoring in data science who are members of the Kaggle platform.
The students should have at least 1 year of programming experience and 1 year of experience with data science tools.

Step 3(b)

5 questions will be given to participants in the form of a 5-Point Likert Scale. The ordinal data collected will help me to rank each individual. For example, a data scientist who chooses (5) for the item (b) means that he/she puts more effort in finding inspirations from other domains in solving existing data science problem compare to data scientists who choose (4), (3), (2) and (1) as their responses. Likewise, a data scientist who chooses (4) puts a bit more effort into exploring this way of solving a problem than data scientists who choose (3), (2), (1) as their choices. We can also tally the ordinal data collected. Let's say we split the sample into an undergraduate student group and graduate student group, and then we count how many data scientists from each group choose each of the responses. We can also further split the undergraduate students into freshman, sophomore, junior and senior, and then compare the responses. We can then plot a bar chart to see the distribution of the data. If we want to gain further insights into the data, we can calculate mode, median, and the interquartile range. For the computation of the mode, it's as simple as the response that most data scientists select. For the median, we will re-arrange the data according to the magnitude first, and then choose the single middle number or two middle numbers.

Step 4(a)

a. Throughout the data science contests, a data scientist with high aspirations will dissect the problem at hand and solve it piece by piece.
b. Throughout the data science contests, a data scientist with high aspirations will try to integrate solutions from different domains to solve the existing data science problem.
c. Throughout the Kaggle data science contest, a data scientist with high aspirations will generate many alternatives before selecting a final solution for submission.
d. Throughout the Kaggle data science contest, a data scientist with high aspirations will try to innovate new models instead of implementing models that are generally used by others.
e. Throughout the Kaggle data science contest, a data scientist with high aspirations will also spend considerably more time researching the information that will help them create a new model that best fits the data.

Step 4(b)

All items listed above can be measured within a ‘short’ time duration right after a one-week data science contest is over.

Step 5

5 rating scale survey questionnaire items will be given to participants after the serial data science contest

1. I dissect the entire data science problem into smaller pieces to solve it efficiently.

(5)Strongly Agree (4)Agree (3)Neither Agree Nor Disagree (2) Disagree (1)Strongly Disagree

1. I integrate solutions from different domains to solve the existing data science problem assigned.

(5)Strongly Agree (4)Agree (3)Neither Agree Nor Disagree (2) Disagree (1)Strongly Disagree

1. I generate many alternatives before selecting a final solution for submission.

(5)Strongly Agree (4)Agree (3)Neither Agree Nor Disagree (2) Disagree (1)Strongly Disagree

1. I try to innovate new models instead of using models that are generally used by others to solve the data science problem.

(5)Strongly Agree (4)Agree (3)Neither Agree Nor Disagree (2) Disagree (1)Strongly Disagree

1. I spend a considerable amount of time researching the information that will help me create a new model that best fits the data.

(5)Strongly Agree (4)Agree (3)Neither Agree Nor Disagree (2) Disagree (1)Strongly Disagree

Step 6

For the first question, a data scientist will need to re-think if they ever dissect a data science problem into multiple parts. In this sense, they will need to think if they solve the data science problem step by step to fully understand the problem at hand and not miss out on the minor details that might affect the accuracy and performance of the model that they develop. If they do this for every sentence or almost every sentence that they read, the respondents are likely to choose “strongly agree” or if they do not do this at all, the respondents will pick “strongly disagree”.

For the second question, a data scientist will need to re-think if they ever consider solutions from different domains to solve the data science problem at hand. For example, the data science problem presented to the respondents is to predict agriculture yield, and they utilize concepts from mathematics. In this sense, the respondents have thought about looking at the problem from the perspectives of the other domains. Depending on how strongly they have ever thought about integrating ideas from other domains, the respondents will choose the option accordingly.

For the third question, a data scientist will need to re-think if they ever develop multiple solutions with the end goal of choosing the best solution for submission. If they have done this for almost every minor problem, then the respondents are expected to choose “strongly agree” as their answer; if they have never done this throughout the entire data science contests, then the respondents are expected to choose “strongly disagree”.

For the fourth question, a data scientist will re-think if they ever think out of the box to come up with innovative solutions. For example, instead of going with the solutions that are generally used to solve a particular problem, a data scientist tries to come out with a different model. If there is a significant effort made in this aspect, the respondents will pick “strongly agree”. If the respondents are the type of data scientists who prefer to go with the safer route of using a generally used model at all times, the data scientists would pick “strongly disagree”.

For the fifth question, a data scientist will re-think if they ever spend a considerable amount of time researching the information that will help them in creating a new model that best fits the data. For example, a data scientist would do a literature search on the latest techniques that are used to analyze the specific types of data given. If they have spent substantial time doing this to give them the best information, then the respondents would select “strongly agree”. If the respondents did not do any research before creating the model, then the respondents would select “strongly disagree”.

Step 7

To compute the total score for each respondent, I will sum up the scores that each respondent gets in the following way with the following two assumptions:

The first assumption is that the distance between two neighboring categories is the same. For example, the effort used to go from “agree” to “strongly agree” is the same as the effort to go from “disagree to “neither agree nor disagree”.

The second assumption is that the weight of each item is the same. Item (a), (b), (c), (d), (e) are of the same weights.

If the participants choose

(a) Strongly Agree – 5 points
(b) Agree – 4 points
(c) Neither Agree Nor Disagree – 3 points
(d) Disagree – 2 points

(e) Strongly Disagree – 1 point

Therefore, the total score for participant 1 would be

5 + 5 + 2 + 1 + 3 = 16 points

The total score for participant 2 would be

4 + 5 + 3 + 1 + 1 = 14 points

With the assumptions stated above, we can say that participant 1 has a slightly higher level of aspirations compare to participant 2.

Step 8

Without validation, the scores from my instrument will not mean much. The score computation on step 7 is meant to be a rough guideline towards the level of aspiration for each individual. For example, an individual who scored 25 points will much likely have a higher level of aspirations to succeed in the data science contest compare to his peers who scored 10 points. Further data analysis needs to be done to examine its validity and reliability. The reliability coefficients such as those suggested by Gadermann et al. (2012) need to be applied to determine the reliability of this ordinal data. The authors suggest the use of "ordinal, polychoric correlation-based versions of reliability coefficients" for quantifying the reliability of the item response data (Gadermann et al., n.d.).

As suggested in step 3(b), the use of mode, median and interquartile range might be appropriate to get an insight from this ordinal data. The interquartile range is the result of the deduction of Q1 from Q3 after the data is arranged according to the magnitude. In this sense, Q1 represents the middle value of the lower half of the data, and Q3 represents the middle value of the upper half of the data.

References

Ali, I. (2019). Personality traits, individual innovativeness and satisfaction with life. *Journal of Innovation & Knowledge*, *4*(1), 38–46. https://doi.org/10.1016/j.jik.2017.11.002

Ashby, J. S., & Schoon, I. (2010). Career success: The role of teenage career aspirations, ambition value and gender in predicting adult social status and earnings. *Journal of Vocational Behavior*, *77*(3), 350–360. https://doi.org/10.1016/j.jvb.2010.06.006

Bokan, I., Buljan, I., Marušić, M., Malički, M., & Marušić, A. (2020). Predictors of academic success and aspirations in secondary nursing education: A cross-sectional study in Croatia. *Nurse Education Today*, *88*, 104370. https://doi.org/10.1016/j.nedt.2020.104370

Gadermann, A. M., Guhn, M., & Zumbo, B. D. (n.d.). *Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide*. https://doi.org/10.7275/N560-J767

Graham, C., & Ruiz Pozuelo, J. (2021). *Do High Aspirations Lead to Better Outcomes? Evidence from a Longitudinal Survey of Adolescents in Peru Do High Aspirations Lead to Better Outcomes? Evidence from a Longitudinal Survey of Adolescents in Peru*. https://doi.org/10.13140/RG.2.2.35873.43364

Hirschi, A., & Spurk, D. (2021). Ambitious employees: Why and when ambition relates to performance and organizational commitment. *Journal of Vocational Behavior*, *127*, 103576. https://doi.org/10.1016/j.jvb.2021.103576

Kaggle: Your Machine Learning and Data Science Community. (n.d.). Retrieved May 23, 2021, from https://www.kaggle.com/

Nunnally, J. C. (1978). *Psychometric theory*. McGraw-Hill.

Valeriu, D. (2015). The Aspiration to Success in School and the Need for Performance of the Pupils. *Procedia - Social and Behavioral Sciences*, *180*, 549–553. https://doi.org/10.1016/j.sbspro.2015.02.158

Zámečník, R. (2014). The Measurement of Employee Motivation by Using Multi-factor Statistical Analysis. *Procedia - Social and Behavioral Sciences*, *109*, 851–857. https://doi.org/10.1016/j.sbspro.2013.12.553