**Data Collection Plan for BSE22 -13**

|  |  |  |
| --- | --- | --- |
|  | **Group Number** | **BSE22-13** |
|  | **Project Title** | **Augmented and Automated Insurance Underwriting** |
|  | Objectives of data collection | To determine a broad view on the use of/ interest in automated underwriting, why choose such a system and the level of satisfaction for using such a system. |
|  | Organization/ community to collect data from. Give reason why the organization was selected | 1. **Sanlam Insurance** – It’s Accessibility and it is in line with the type of product we want to research about [ Life and Health Insurance ]  2. **Kaggle** – Because it has a large collection of datasets that have data about insurance Underwriting |
|  | Category of respondents to collect data from. Reasons why the categories were selected | **Insurance Underwriters** – representatives were selected on the basis of their knowledge of the company’s automated underwriting process. These individuals were primarily underwriters, but actuaries, claims managers, and other personnel also can participate in the interview.  **Users (insureds)** – They are the end product of the intented system. |
|  | Data collection methods to be used. Give reasons for your choices. See [https://www.scribbr.com/methodology/data-collection/#step-2-choose-your-data-collection-method](https://www.scribbr.com/methodology/data-collection/" \l "step-2-choose-your-data-collection-method) | **Interview** – Interviews provide the best method of developing a deeper understanding of subject matter (Automated Underwriting )  **Survey –** using google forms (Questionnaires) to know what people think about insurance underwriting and how it can be improved through automation.  **Secondary Data Collection** – We need to study existing datasets on insurance and know the variables that are used to determine eligibility for a certain insurance policy.  **Chosen Dataset** is [Prudential Life Insurance Dataset on Kaggle](https://www.kaggle.com/c/prudential-life-insurance-assessment/data) |
|  | Sample sizes of the respondents per type of method to use. | 30 – insurance underwriters  20 - users |
|  | Data collection administration methods. Reason | Interviews, Questionnaires, literature reviews |
|  | Will you do any pilot for your data collection tools? Which category of people will you use? | Yes, Insurance Underwriters |
|  | Will you need an introductory letter from the department? | Yes. |
|  | What kinds of data do you plan to collect and to what relevance will it be to the design of your system? | - Parameters that are used to asses risk for a particular insurance application.  - The process of underwriting. |
|  | Do you plan to return with some records from the data collection process? If so, which ones? | - A recording from an Interview,  - Notes from interview/focus groups  - |
|  | How do you intend to analyse the data?  Read this  <https://sixsigmastudyguide.com/types-of-data/> | Exploratory Data Analysis using Visual Analytic [ bar charts, heat-maps, correlation tables e.t.c.]  Encoding data and attaching weights in order to determine to quantify the data and get relationships between the different variables. |
|  | Artefacts to bring from the field (eg. Recordings (**Ask for permission if you want to record**), images, receipts, registration forms etc) | - Recording, Notes, Dataset, filled Forms. |