COMP0035 2023/24 Coursework 1

*1. Data preparation and understanding­*

*Short summary*

*1.4 Explanation of code preparation & understanding*

The data set presents information registered by the Dog Control Department of the Zurich City Police on the dog populations in the city of Zurich since 2015 to 2023. This involves data on the dogs, their owners and their municipal registration. Information on the owner is given on the age group, gender and statistical area of ​​residence, similarly, the dog information recorded is based on the breed, breed type, gender, year of birth, age and colour. The attributes of this data allows for possible statistical distributions since the data is prearranged over the period from 2015 to 2023. Additionally, relationships between different variables could be investigated to predict potential occurrences.

The factors are then described using three different methods; long method which includes a description, code method that is established by using a numbering system which simplifies the future analysis and exploration of the data and lastly, sorting method which is based on the attributes considered using a system of categorization which also aids the future analysis of the data.

In order to prepare the code, I decided to reduce some of the sections of the data to code method and long method for potential use while simplifying the implementation and execution. Additionally, taking into account the data is only available in German, I had some issues interpreting and understanding the data at first glance and required the use of a translation tool. Therefore I decided to add a new header to all of the columns with a translation to English which will grant me simpler manipulation and use in the future. Additionally, in the column of ‘Age of dog owner’ there are unknown values which I decided to remove from the data to ensure consistency in the interpretation.

Consequently, I proceeded to prepare the data in VScode python using the pandas library. Firstly, the data file containing information on the Dog populations in the city of Zurich since 2015 to 2023 was opened and read. Then, the required columns for the analysis and interpretation were selected and appended to a new file (data\_preparation.csv). And lastly, the headers were replaced with an English description of the data.

The data manipulation and interpretation was directed towards dog breeders established in Zurich. The exploration involves the following tabulated attributes:

|  |  |  |  |
| --- | --- | --- | --- |
| Year | Age of dog owner | Gender of dog owner | City district of dog owner |
| Numeric | Numeric | Numeric | Description |
| The year the municipal dog register took place | The age of the dog owner rounded to a multiple of 10 | The gender of the dog owner represented as ‘1’ for male and ‘2’ for female | The city district the dog owner lives in |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Primary breed of dog | Mixed breed | Year of birth of the dog | Age of the dog | Gender of the dog |
| Description | Description | Numeric | Numeric | Numeric |
| The primary breed of the dog registered | Whether the dog is mixed breed, pedigree or unknown | The year of birth of the dog | The age of the dog | The gender of the dog |

*2. Product and project definition*

The focus of the project is Zurich’s dog breeding strategy and it involves comprehensive data on the dog populations in the city of Zurich since 2015 to 2023. The web application, App2, aims to modernize and facilitate how dog breeders operate in Zurich by including a collection of visualizations focused on the needs and objectives of Zurich's dog breeding setting. App2 intends to provide dog breeders in Zurich insights of the market, inspire breeders and smoothen their planning. Overall, the aim of the web application is to encourage breeder to make data-driven decisions, regarding breed focus, the marketing strategies involves and fulfil engagement with the customers.

*2.2 Product overview*

App1: REST API for Developers

The REST API is a product designed for software developers since it allows them to have programmatic access to a unified dataset about Zurich's dog breeding demographics. Making use of the REST API, software developers are encouraged to incorporate, explore and employ this dataset for other intended purposes and applications.

App2: Data-Driven Web Application Product for Breeders in Zurich

App2 is a web application which influences the REST API stipulated by App1, as described above. The web application emphases on aiding dog breeders based in Zurich. The leading qualities of App2 are on insightful data visualizations including dashboards which offer dog breeders to expand their knowledge on the most prevalent dog breeds for each city district in Zurich and how the number of pedigree dogs has changed over time (2015 to 2023), the distribution of pedigree dogs, the distribution of female pedigree dogs, the distribution of male pedigree dogs and the age of the pedigree dogs. This user-friendly interface encourages breeders to engage with valuable insights for potential decision-making-

*2.3 Persona*

Persona for App2: Zurich Dog Breeders

Name: Maria Garcia

Age: 34

Occupation: Professional Dog Breeder in Zurich

Background: Maria Garcia is a dog breeder based in Zurich with experience in the sector as she has subsisted in the industry for over 5 years. Consequently, she has profound understanding on dog pedigrees. Maria is passionate about her career, her purpose as a professional dog breeder, ambitious about endorsing healthy bloodlines and taking on board responsible breeding practices.

Goals, Objectives and Needs:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Data-Driven Decisions | Insight on dog pedigree | Focus on female/male pedigree | User-Friendly Interface | Relevance to local region | Visual representation and understanding |
| Maria is interested in having access to reliable data on the variety of dog breeds in the city districts in Zurich. This allows her to make informed decisions on the focus of dog breed for her breeding program and profession. | Maria is interested in the number of pedigree dogs found in the different city districts in Zurich. Allowing her to understand the market and demand for pedigree dogs and proceeding to tailor the market and breeding strategy appropriately. | Maria has to focus on the male and female distribution of pedigree dogs which allows her to fulfil her breeding strategy appropriately taking into account the market trends. | Maria is proficient in the use of modern technology although prefers to have to work with an intuitive interface. This allows her to access an understand data avoiding technical complications. | Since Maria is a dog breeder based in Zurich she has to takes into account data collected for a specific region. This permits her to have a potential understanding on the possible impacts in her breeding strategy in Zurich. | Maria appreciates the data being accessible in a visual format including graphs since this facilitates comprehension and finding trends within the data quickly. |

Overall, by focusing on fulfilling the needs of dog breeders in Zurich App2’s objective is to deliver and propose a tool which allows breeders to make well-informed decisions.

*3. Tools & techniques*

*3.1 Source code control*

*3.4 Use of AI*

The use of AI for data preparation can result to be an innovative and effective method when handling the data before analysing it.

|  |  |  |
| --- | --- | --- |
| Potential feature ideas | Detection of outliers | Language translation |
| Preparing data and analysis by discussing possible applications and features with ChatGPT. This allowed me to have a wider understanding and applying the data to real-life scenarios | Making queries about my dataset allowed me to find potential outliers and anomalies of the dataset. ChatGPT put the dataset into perspective which permitted me to look into future applications of the dataset and its analysis by avoiding potential errors. | Since my dataset is in German I had to perform abundant translations in order to understand the information available for potential analysis. ChatGPT facilitated the task of translating and smoothened the procedure of indulging the context of the dataset. |

The use of AI overall aided my understanding of the dataset although it was vital to critically evaluate the content provided by ChatGPT.

*5. References*