### 

### **COMP90106 Data Science Project Pt1 Group 12 Written Report**

**Haitong Gao**   **Yuexin Li**

1068254 959634  
 [haitongg@student.unimelb.edu.au](mailto:haitongg@student.unimelb.edu.au) yuexin1@student.unimelb.edu.au

**Eric Luanzon Meijun Yue**

1218692 1190161

[mluanzon@student.unimelb.edu.au](mailto:mluanzon@student.unimelb.edu.au) [meijuny@student.unimelb.edu.au](mailto:meijuny@student.unimelb.edu.au)

**Chi Zhang**

1067750

chzhang1@student.unimelb.edu.au

May 27, 2022  
  
**Abstract**

The accuracy of small-area population forecasts is an important challenge in current society as it provides recommendations and instructions for the community or government to plan future infrastructure facilities in specific areas. The purpose of this capstone project is to develop a machine learning model that helps to provide accurate small area population forecasts. In this report, the introduction of the topic and difficulties from a data science perspective are presented, following a discussion of previous papers and models that provide the potential related solution to the target issue. Then, the data analysis of the provided datasets, including time-series data of the of the total populations for 351 Australian small area populations at the Statistical Area 3 (SA3) level and an additional dataset which provides the populations by age and sex for these small area, are presented in ways of tables and plots. Based on the preprocessed dataset and patterns of population trend observed from each specific small region, preliminary models based on time series forecasting are proposed. At the end of this report, a proposal for this project in Semester 2 is provided, and a timeline table is presented to illustrate the whole working process and key tasks (milestones).  
 *Keywords:* Population Forecasting, Time Series, Machine Learning, Proposal, Timeline