



SUBMISSION FORM

All submission forms must include the following information. Separate submission forms must be turned in for each eligible program. **Deadline: July 1, 2022.** Please include this submission form with the electronic entry. If you do not receive an email confirming receipt of your entry within 3 days of submission, please contact [Gage Harter](#).

PROGRAM INFORMATION

County: _____

Program Title: _____

Program Category: _____

CONTACT INFORMATION

Name: _____

Title: _____

Department: _____

Telephone: _____ Website: _____

Email: _____

SIGNATURE OF COUNTY ADMINISTRATOR OR DEPUTY/ASSISTANT COUNTY ADMINISTRATOR

Name: _____

Title: _____

Signature: _____ 

VACo Award Application

Submission Category

Information Technology: Illustrate how counties can leverage data assets creatively to produce solutions that benefit community design and development processes.

Business Challenge

Chesterfield County's steady population growth generates challenging demands for planning future civic infrastructure and service needs. Planning leaders often depend on local knowledge and intuition to influence ideas. Ideas do not always turn out the way they intended, which can lead to expensive mistakes in retrospect. Our commitment to leveraging our data assets for decisions making led us to designing a data-driven approach to generate predictions. These predictions supplement the traditional intuitive leadership approach to provide improved decisions for the best ways to invest taxpayer revenues back into the community.

Brief Overview

The Community Forecasting project blends census demographics with real estate housing trends to predict future service needs for our residents, businesses and visitors. Aligned with the Capital Improvement Program (CIP), the team evaluated forecasts that influence Parks and Recreation (P&R) and Chesterfield County Public Library (CCPL) decision making. The cross functional team included Deputy County Administrators; directors from Budget and Management, Parks and Recreation, and CCPL; subject matter experts; technology leaders and staff; data engineers and data scientists; and help from Catapult Systems, a Microsoft Gold Partner.

Previously, the Comprehensive Planning unit had forecasted population at a county-at-large basis. We began with P&R and CCPL data sources. The team quickly realized that the quality and completeness of existing departmental data was not sufficient for machine learning algorithms to predict future outcomes. In addition, the University of Virginia Weldon Cooper Center for Public Service Demographics Research Group published quality concerns regarding 2020 census data for census blocks. Since service delivery for both organizations depend on population trends, the team shifted to predicting population by leveraging Chesterfield real estate data assets combined with census demographic data. Our approach employs scientific models that consume custom generated geographical clusters, grounded in number of bedrooms by housing type. This approach was reviewed with the UVA Weldon Cooper Center for Public Service Demographics Research Group for validation.

Executive Summary

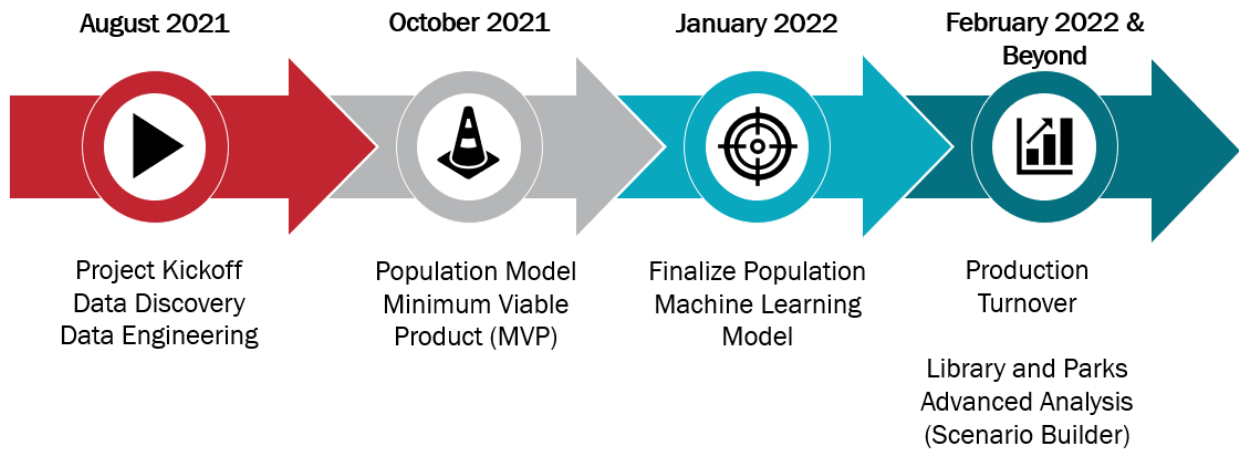
Project Name

Community Forecasting: Data-Driven Population Projections

Program Objective

Develop data-driven insights and forecasts that form the basis for future decisions on capital expenses for Parks and Recreation (P&R) and Chesterfield County Public Library (CCPL).

Time Frame



- **August 2021:** Project Kickoff, Data Discovery, Data Engineering
- **October 2021:** Population Model, Minimum Viable Product (MVP)
- **January 2022:** Finalize Population Machine Learning Model
- **February 2022 and Beyond:** Production Turnover, CCPL and P&R Advanced Analysis (Scenario Builder)

Involvement

The project was initiated with direct sponsorship and involvement from County Administration, Parks and Recreation and Chesterfield County Public Library. As the project shifted to a population-based approach, sponsorship shifted directly to our Budget and Management Department which provided a core level understanding for delivering our county services. The P&R and CCPL sponsors, and other domain experts, remained engaged through the modeling phase of the project. They then became responsible for additional analysis of core P&R-specific or CCPL-specific data to understand the unique needs for their respective service delivery areas. The result was a successfully innovative collaboration between Information Services Technology (IST), Parks and Recreation, Chesterfield County Public Libraries, and Budget and Management, with help from Catapult Systems.

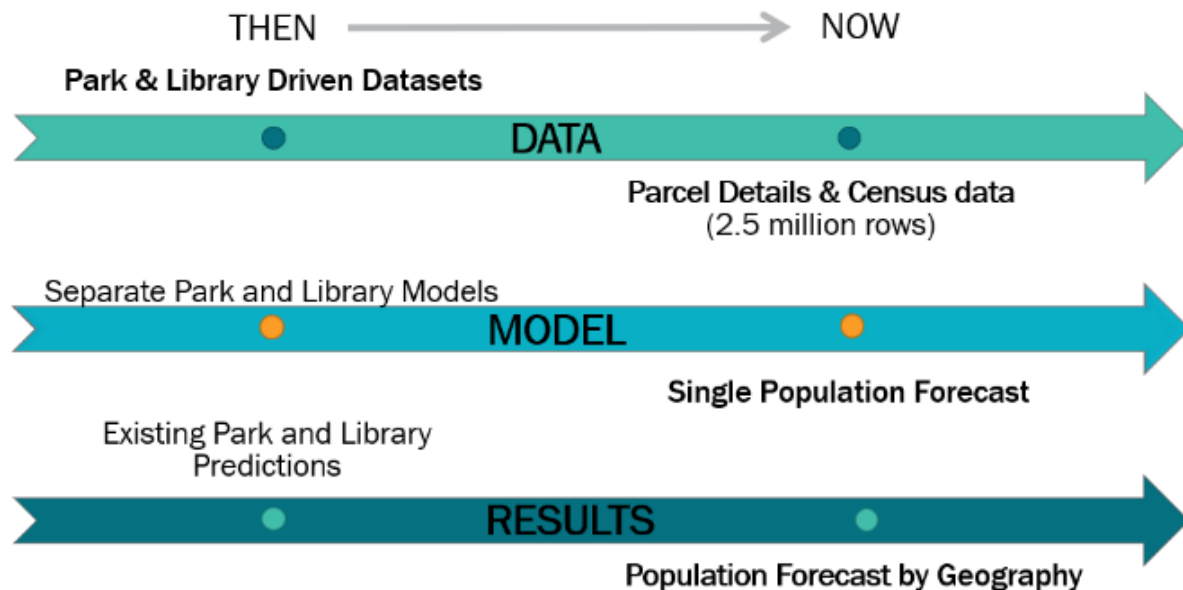
Project Scope

- Conduct exploratory data analysis for P&R and CCPL with a focus on the geographic dispersion of services to reaching all county resident populations.
- Prioritize contributing factors for the data science approach to ensure limited data engineering requirements for delivering the minimum viable product.
- Provision Microsoft Azure cloud-based resources to support the tools and technologies associated with the data engineering and modelling effort.
- Settle on the total number of data sources that will be used to create new pipelines in line with the MVP contributing factors. These will be brought into the county's community data sources stored in Azure. Execute data sharing agreements where necessary. Classify any sensitive data.
- Perform scientific experimentation with models and propose best-fit option(s). There will be a fixed number of sprints that will accommodate as many experiments as possible within the time period.

- Deliver production ready population projection forecasting model for county budget to use in production to form the basis of analysis specific to parks and libraries and other county services.
- Deliver analytical data sets combining population model outputs and P&R and/or CCPL trends that were collected into the community warehouse.
- Review results with the team and stakeholders continuously through the Kanban approach. Provide a release demo for the project sponsors through a weekly meeting.
- Effect knowledge transfer via project status, demos and team collaboration in keeping with Kanban principles (in lieu of formal training).
- Provide engineering support for Lakehouse Architecture. Identify potential plans to refactor our current architecture.

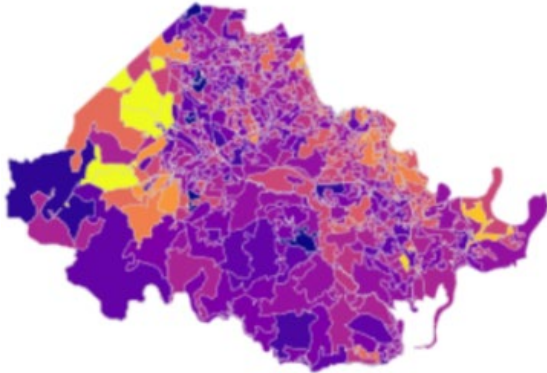
Major Project Shift

During the data discovery phase of the project, two key concepts emerged. First, the quality and completeness of the core departmental data for both P&R and CCPL was insufficient for a predictive modeling approach. Second, the themes regarding delivering services for both departments had similar needs to meeting the needs of the county constituents and was based on where the people will be in the future. Thus, the team shifted to our population specific approach.



Ideation and validation of approach was completed with the University of Virginia Weldon Cooper Center for Public Service Demographics Research Group leaders. The introduction of a population proxy grounded in bedrooms by housing type, provided by our extremely reliable 20+ year history of real estate data, provided the key components to move our modeling approach forward.

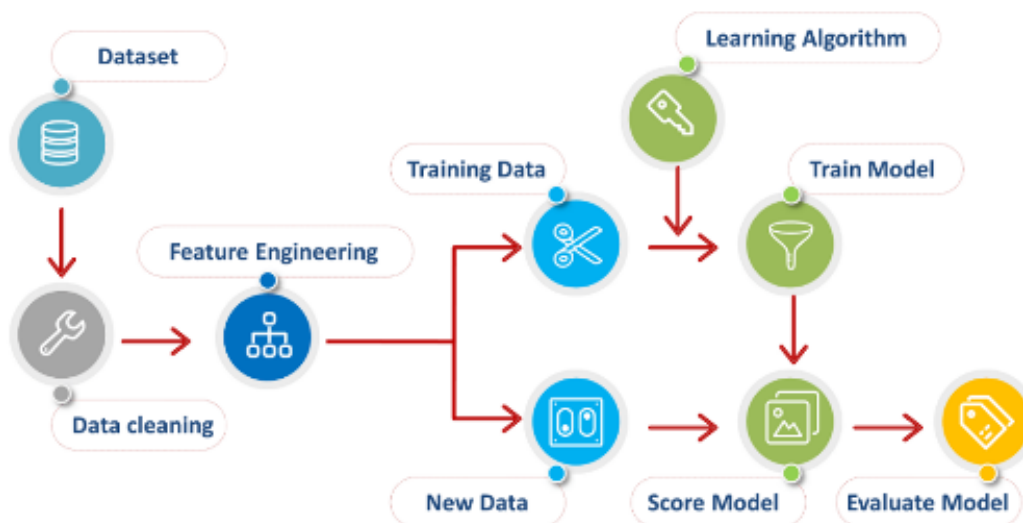
Create Custom Community Geographic Clusters



- Procedurally generate homogenous areas
- Areas border each other
- Each cluster contains at least 450 people
- Custom geographies are the combinations of census blocks
- Generated using housing data similarity of the census blocks

Fundamental to our scientific approach is the ability to create custom community based geographic clusters, in which we aggregate our data. These clusters are optimally sized areas of similar characteristics that significantly improve the accuracy of the results. This enables a reliable baseline of existing growth to support future growth prediction. The clusters are where the team marries census and housing data together.

Modeling Process



The data scientists from Catapult used this rich dataset for the experimentation and modeling process which optimizes features and algorithms used to train and evaluate optimum model performance. This activity went through several iterations between October 2021 and January 2022 when until the production-ready model was completed.

After the population forecasting model was completed, emphasis shifted to enabling Budget and Management, P&R and CCPL team members with tools and support to enhance forecasts with domain specific datasets which further refine predicted needs for county services delivered by their departments.

During the initial discovery phase of the project, P&R and CCPL staff shared innovative ideas regarding the future distribution and equitable accessibility of their services. Each department has goals for accessibility by driving, biking or walking. Generally speaking, services should be within a 15-minute walk, 10-minute bike or a 5-minute drive. These distances are being used in advanced analysis regarding future projects for each department. Additional data is included for broadband availability which influences literacy rates according to the Virginia Department of Education.

Program Cost

Our consulting costs for Catapult Data scientists, engineers and project manager were \$397,633.80. The staff costs were a core team of six county employees spending from 8-24 hours weekly for the eight-month duration of the project. Operational costs for Microsoft Azure resources for the project over the eight months was \$1,511.97. Automation with Azure DevOps has been implemented that ensures Azure resources are removed except for the annual execution of the five-year forecast model or experimentation from the county staff managing the model. During the month of March 2022 when several iterations of the production turnover were being tested, the total cost was under \$600. Investment in Azure cloud services would be required for another locality wanting to duplicate our approach. Resources deployed include Azure Data Factory for data movement and processing, Azure SQL Database/Azure Data Lake Storage for data storage, and Azure Machine Learning for integration of Python scripts and open-source machine learning libraries. Azure DevOps is used for source control and continuous integration/continuous deployment processes.

Awards Criteria Fulfillment

Innovation

The project combined census data with detailed real estate data to produce a more accurate result and a product that can be leveraged by other departments. In addition, we applied algorithms to manufacture regions that add value to the predictive modeling process.

Partnering or Collaboration

This was a collaborative effort between Information Services Technology, Parks and Recreation, Chesterfield County Public Library, and Budget and Management, with input and review from UVA Weldon Cooper Center for Public Service Demographics Research Group which we aligned with a partnership between Chesterfield County and Catapult Systems.

Model for Other Localities

Our approach could easily be implemented by any locality that has historical detailed real estate data to leverage.