

## Introduction

Avaya Research Center is evaluating the housing market in the Greater Toronto Area using third-party datasets collected by external sources. This information will provide a better understanding of what drives housing prices. The purpose of this project is to find fact based insights on meaningful patterns in the housing market. The stakes are high in the real estate industry. If potential home buyers evaluate their purchasing correctly they can have a bigger return on their investment when it's time to resell. Make a bad decisions by not getting to know the neighbourhood history before purchasing can lead to a lost when trying to resell. With improved historic data decisions can be made based on facts and data collected. In the past, home builders relied on their gut and focus groups. Avaya was able to obtain a dataset containing house sale prices from Listing.Ca, Toronto Police Service (Public Safety Data Portal), and Simply Analytics (Income) for the surrounding Toronto area.

## Objectives

Determine factors that will influence home buyers the most for example crimes in the neighbourhood and their household income. Within the market, different amenities appeal to different consumer groups. Some consumers are first time buyers, they prioritize space for growing families, others are empty nesters looking to downsize. Everyone still wants their money to go further. The overall objective is to determine how does crime in neighborhoods affects housing cost other factors that affects or drives housing cost.

## Problem Statement

Focusing on the areas that has the most crime activities and determining the correlation between the cost of the houses in those areas and the household income and try to factor in the level of education in those areas as well.

## Data Sources

The datasets were gathered from several different third party websitesuch as the Toronto Public Safety Data Portal, Listing.Ca, Simply Analytics and.

## Collection Method

## Analysis Methods

## Timeline

<b>Week 1:</b>	-Signup for GitHub, research python libraries and data APIs. Start developing a topic area/dataset for inquiry.
<b>Week 2:</b>	-Begin data collection, assembling methodology for Exploratory Data

	<p>Analysis.</p> <ul style="list-style-type: none"> <li>-ETL work into data warehousing</li> </ul> <p>Begin exploring data and documenting issues/limitations/needs understanding needs/limitations regarding research question (might need to adjust question, scope, data, etc.).</p>
<b>Week 3:</b>	<ul style="list-style-type: none"> <li>-Submission of EDA sprint (via GitHub/Moodle).</li> <li>-Collect/Augment/Refine project according to Sprint 1 findings.</li> <li>-Develop several analytical methodologies (these should be methods you are interested in learning – like Topic modelling or sentiment analysis, or social network analysis, etc.)</li> </ul>
<b>Week 4:</b>	<ul style="list-style-type: none"> <li>-All data collection should be completed</li> <li>The methodology should be finalized</li> </ul>
<b>Week 5</b>	<ul style="list-style-type: none"> <li>-Submission of Sprint 2 to GitHub/. Includes codebase, report (brief), and plan for analysis</li> </ul>
<b>Week 6</b>	<ul style="list-style-type: none"> <li>-Start final sprint prepare for final submission.</li> </ul>
<b>Week 7</b>	<ul style="list-style-type: none"> <li>-Work on final project.</li> </ul>
<b>Week 8</b>	<p>Final report – pdf/word doc</p> <ul style="list-style-type: none"> <li>-3-5 pages of background, research question, methodology, results, conclusion</li> <li>-GitHub Repo</li> <li>-Well documented – meaning someone who comes across it randomly should know exactly what it is, how to re-create it, and what your repo contains</li> <li>-Final notebooks/analysis/results</li> <li>-Each sprint notebook documented,</li> </ul>

	detailed, and refined.
--	------------------------

### **Document Overview**

The remainder of this document details the requirements, technical approach, expected results and project plan. This sections will describe the plan on providing a solution to the problem statement, and achieve the scope.