**VISUALISING CAMBRIDGE PROPERTY DATABASE**

**USING QLIK**

Logo

Description automatically generated

**ALY6070 - Communication and Visualization for Data Analytics**

**Instructed by Prof. Jack Bergersen**

**Submitted on May 6, 2022**

**Team 5:**

Dhamodara Venkata Ganesha Bhar Pecchetti

Perumal Kabali

Sudhamshu Vidyananda

Teerajate Vantanasiri

Vidhya Lakshmi Vaithilingam Palanimurugan

**INTRODUCTION**

Qlik is a cloud-based analytical tool widely used to create interactive data visualization. In this Qlik exploration assignment, we have analyzed the Cambridge Property Database dataset. A Cambridge Property FY2016 to-FY2021 dataset captures real estate details related to property in Cambridge, Massachusetts. This dataset incorporates a wide variety of information on many properties of a real estate asset, some of them are land area, LandValue, and type and class of properties in Cambridge, the dataset even includes, book/page entry of the property, sale date, property particulars such as types and number of rooms exterior and interior materials description which facilitated us to perform various analysis. It provides valuations calculated by the Cambridge Valuation Department for each property from 2016 to 2021. The sum of "Land Value" and "Building Value" determines the evaluation value. In 2021, the asking price for all Cambridge homes is $ 2,900,000 million. The price quoted for the lot in the final certificate transfer is provided by the SalePrice attribute. To capture its coordinates on the map, each property is recognized by its unique parcel ID and GIS (Geographic Information System) ID, as well as the longitude and latitude of the location. Properties are organized into 119 categories covering residential, commercial, and public spaces based on the PropertyClass attribute. The ResidentialExemption attribute contains information about the properties that the owner occupies. The PropertyTaxAmount attribute contains information about the tax levied on each property in a particular fiscal year. Cambridge offers one of the state's most generous housing tax exemptions. Cambridge is divided into 29 tax districts based on their geographical location.

**PROBLEM STATEMENTS**

Many research questions can be asked and answered with the Cambridge Property Database. This project answers the below two questions raised on the given dataset.

1. What are the top 10 Tax Districts by the tax revenue collected? What property category is contributing the highest to tax payment (residential or commercial)? What are the property types of these high contributors?
2. What are the top 10 Residential Property Categories in Cambridge which have high Assessed Property Value? Does their value change over the years? Is there a trend?

**DASHBOARD 1**

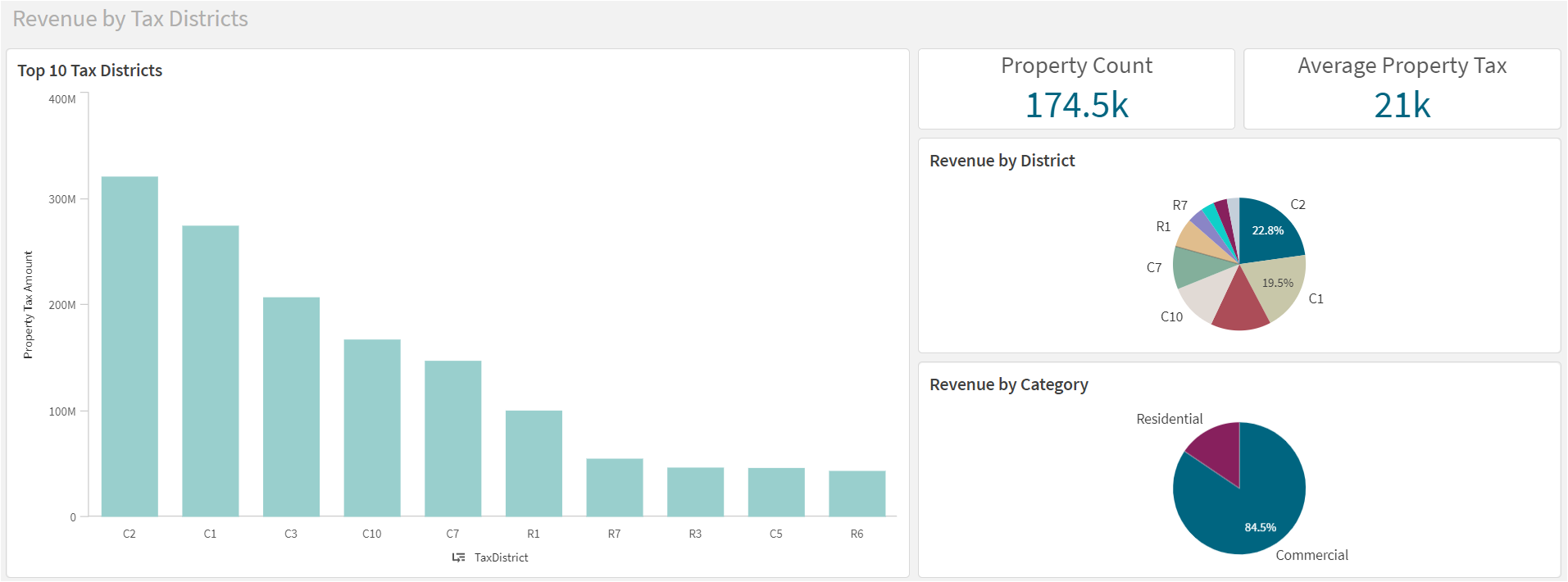
To answer the first research question, we have created a dashboard with a **Bar graph**, **Pie charts,** and **two KPIs**.

**Variables involved:**

* Tax District
* Property Tax Amount
* Residential Exemption
* Property Class

All properties which have a Residential Exemption are categorized as “Residential” and the others as “Commercial”.

**Method of Analysis:** Exploratory Data Analysis



*Dashboard 1: Tax collected in Cambridge analyzed by Tax Districts*

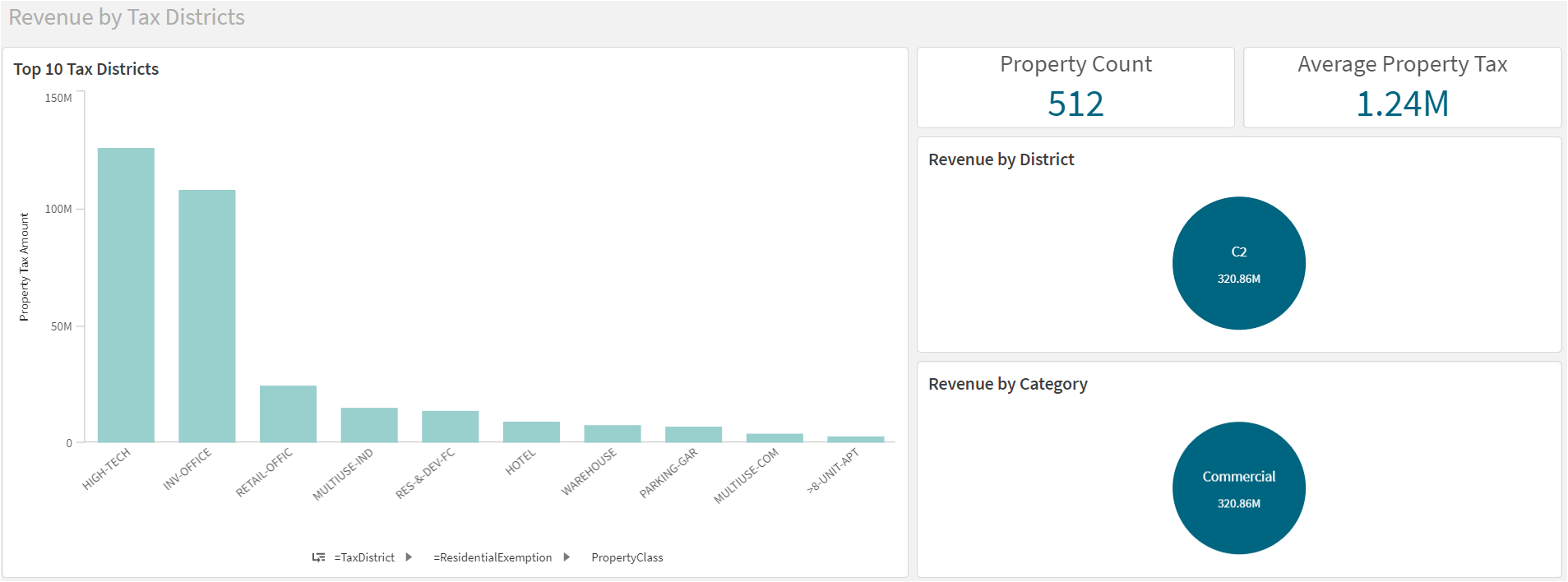
**Design Decisions and Dashboard Summary:**

**Bar Graph**

**A bar graph is used to display the top 10 Tax Districts by the total tax amount collected.** We have used the **drill-down** feature of Qlik to show a deeper picture of what property class belongs to those tax districts and if they are exempted from the residential exemption option.

The reason for choosing the bar graph and the drill-down method is that the bar graph is usually used when we want to compare data between 2 or more categories. It is straightforward to understand and visually looks neat and clean. Drill down feature in Qlik allows users to move from a high-level overview to a more thorough and granular view of the same dataset they are looking at just by clicking on any required measure in the graph.

The top Tax Districts are plotted along the x-axis and the sum of Property Tax amounts for those tax districts is plotted along the y-axis. We see that **C2** has the highest property tax **(320.86M)** followed by **C1 (274.55M)** and **C3 (207.03M).** These are clearly differentiated by the size of the bars and the sum amount can be viewed by hovering around the bars for each Tax district. If we want to check the property classes that fall under the tax district, we can select the required tax district bar and it automatically shows the property class under the tax district and display the total tax amount paid by that particular property class. For the C2 tax districts, the highest tax amount is paid by **HIGH-TECH property (125.79M**) followed by **INV-OFFICE (107.9M)** and **RETAIL-OFFICE (24.35M)**.



*Figure 1:Tax Collected in C2 Tax District*

**KPIs**

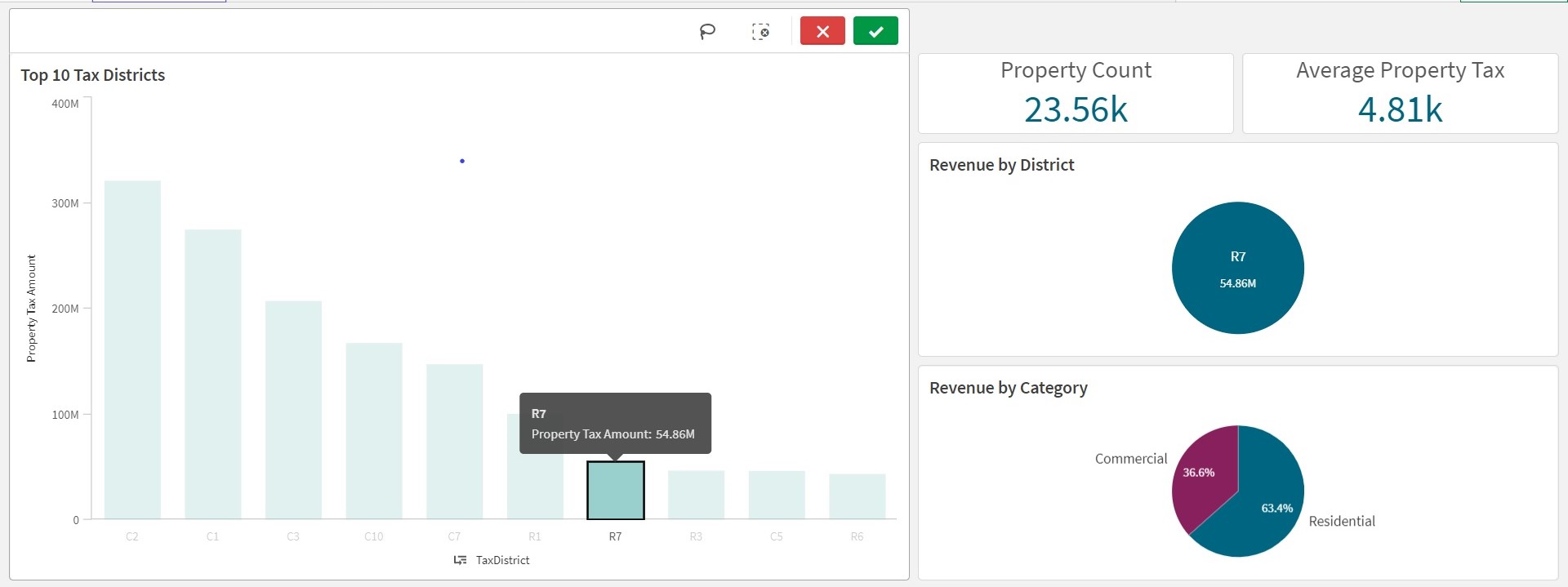
**KPIs are used to dynamically show the total property count and average property tax paid for each tax district and property class**. We have opted for these KPIs to dynamically show the property counts and average tax amount directly to the user which makes it user-friendly, and they need not struggle to find the value. It is in the prime location in the dashboard above the pie charts so that whenever users select the data they are interested in, they can spontaneously see the values.

For the top 10 tax districts, the total property count is **174.5k** and the average property tax paid is **21k**. If we select any tax district bar, for example, C2, we see that the property count gets automatically updated to **512** and the average tax amount to **1.24M**. This informs that there are a total of **512** property classes in **C2** tax districts based on different attributes and the average tax amount paid by the properties is **1.24M**. If we drill down deeper by selecting the property that we are interested in, for example, **HIGH–TECH**, we see that the property count is **79**, which states that there are **79** different **HIGH-TECH** properties in the **C2** district and the average tax amount paid by this property is **3.07M**.

**Pie Charts**

**We have used two pie charts on the right side of the dashboard which show Revenue by Tax District and Revenue by Category.** Pie charts show the data comparison briefly to analyze the data effectively and quickly. We have opted for pie charts since we wanted to show the percentage value of the attributes to get a deeper understanding that will boast the dashboard content we are putting up.

**Revenue by District** pie chart shows the percentage of tax paid by the top 10 tax districts. **C2** has the highest percentage of **22.8** followed by **C1 (19.5%)** and **C3 (14.7%).** This is evident because the bar graph also shows that C2 is having the highest sum of property tax followed by C1 and C3. This chart is also dynamic, and we can select the tax district that we are interested in by selecting the required slice. Each slice corresponds to tax districts and they are differentiated by color. If we select the required slice, we see the total property tax amount along with the percentage value for that tax district.



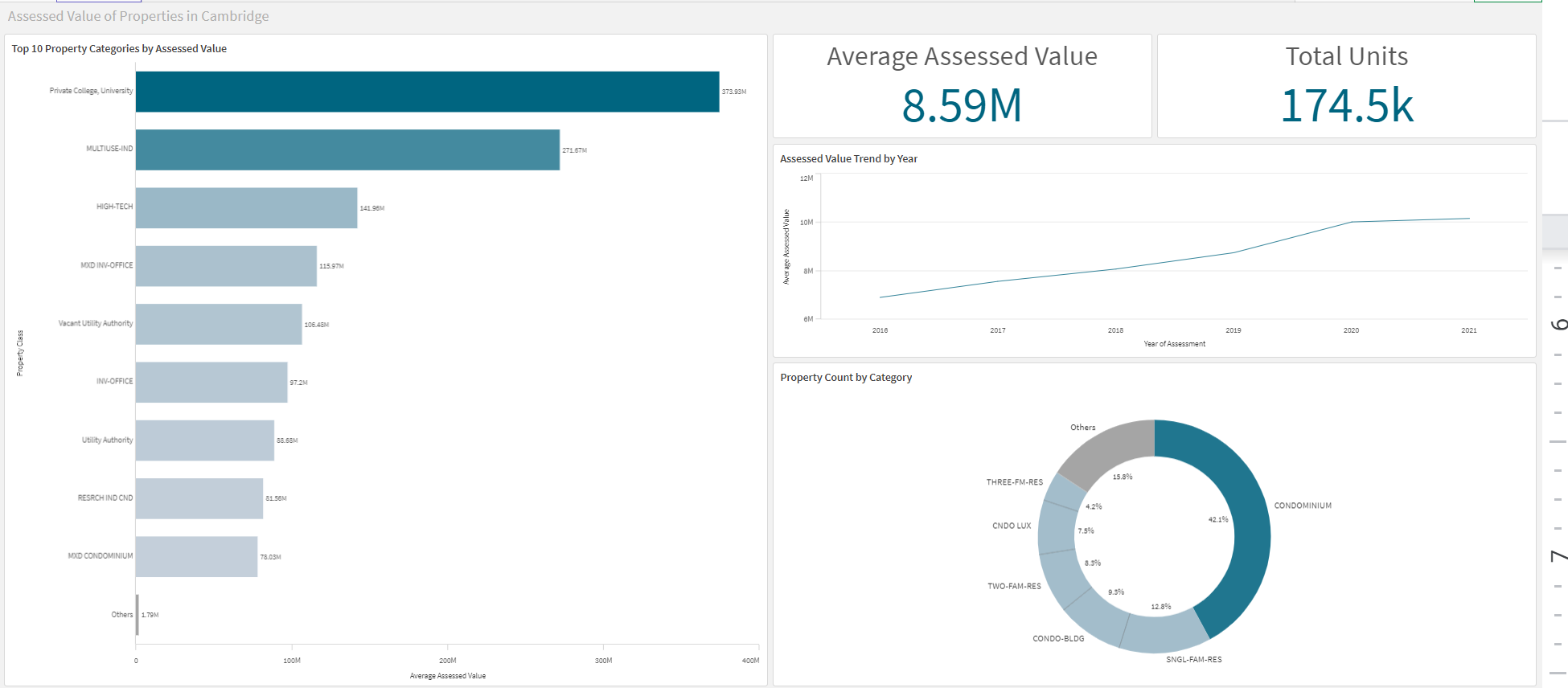
*Figure 2:Tax Collected in R7 Tax District*

**Revenue by Category** pie chart shows the percentage of properties categorized by residential and commercial filters based on residential exemption attributes. It is evident from the pie chart that the commercial sector is having the highest revenue of **84.5%.** This informs that there are more commercial properties than residential ones. Residential fills up to only **15.5%** of the revenue generated. We can hover over each slice to get more details on the attributes. **From figure 2,** we can say that **Tax District R7 has more tax from Residential properties (63.4%) than Commercial properties (36.6%).**

An important feature of this dashboard is that it is **interactive**, and users can play with the data between bar and pie charts. If we select a tax district in the bar graph, the data gets automatically **updated in both the pie charts**. For example, if we select the C2 tax district in the bar graph, the revenue by district pie chart shows the total property tax amount generated in the C2 tax district, and the revenue by category pie chart shows the percentage of commercial/residential property type in the C2 district. It is evident that there are only commercial properties in the C2 district and shows the total property amount tax generated. This will help users to see the data directly instead of digging deeper through the values in the dashboard and makes the dashboard user-friendly.

**DASHBOARD 2**

This interactive dashboard analyzes the average assessed values of Cambridge's top 10 property classes. **The dashboard is built with three charts (a Bar chart, a line chart, and a Donut chart) and two KPIs** and answers the proposed second research question.

****

*Dashboard 2: Analyzing assessed values of properties in Cambridge*

**Variables involved:**

* Tax District
* Assessed Value
* Residential Exemption
* Property Class
* PID

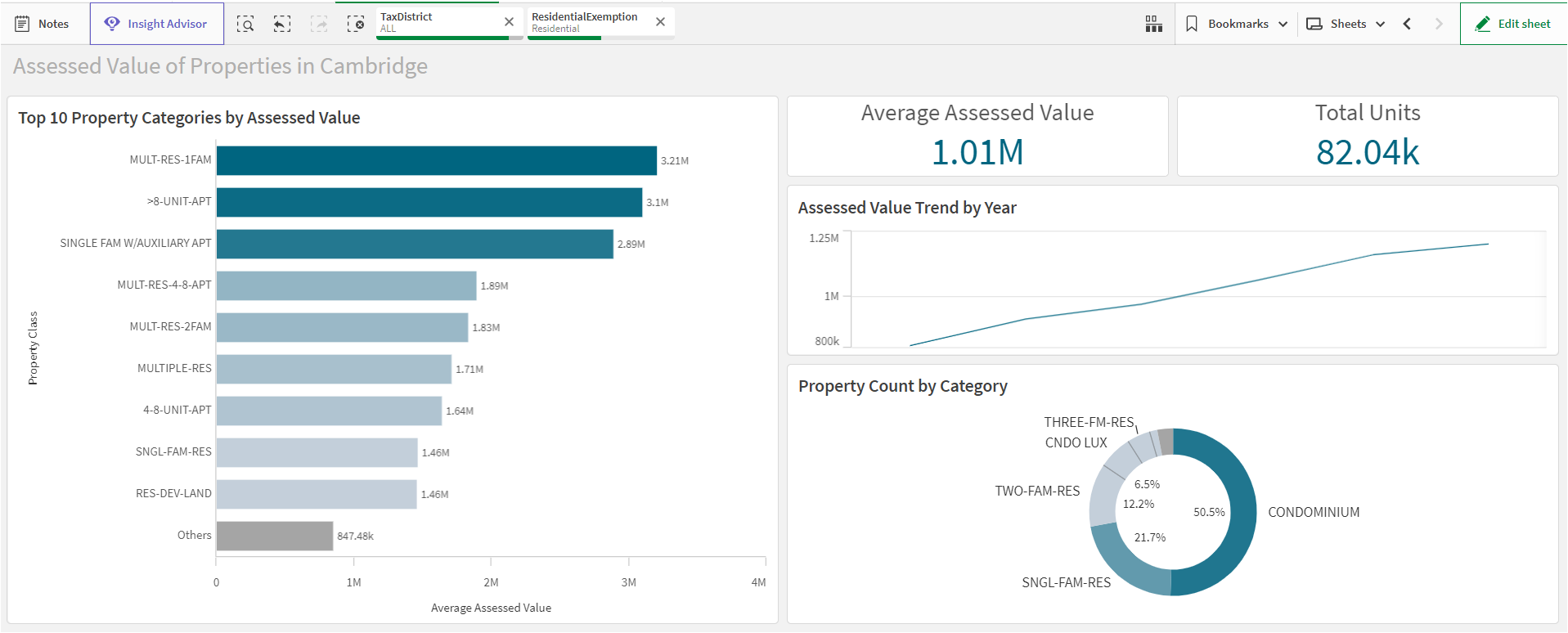
All properties which have a Residential Exemption are categorized as “Residential” and the others as “Commercial”.

**Method of Analysis:** Exploratory Data Analysis

**Dashboard Summary:**

**Bar Graph**

**The Bar graph located on the left side of the dashboard displays the top ten properties by property class.** We have chosen horizontal style bars and sorted the values in descending order to display the highest values on top. To achieve that, the X-axis represents the property class, and the Y-axis represents the average assessed value in dollars ranging from **0 to 400 million** US dollars. The value of the property class is displayed on each graph to know the value instantly. From the bar chart, it is evident that the private colleges and Universities' average assessed value is approximately **$374 million**, followed by **$271.67** million for multi-purpose buildings, and the lowest average values are observed at **$78.03** million and **$1.79** million for the mixed condominium and others respectively. When we filter the residential properties using ResidentialExemption, we can see that **“MULT-RES-1FAM”**  has the highest assessed value of **3.21 million** US dollars as shown in figure 3.



*Figure 3: Assessed Value of Residential properties across Cambridge*

**KPIs**

Two KPIs are placed on the top right corner of the dashboard. One shows the **average assessed value** of selection and the other **total units(distinct properties)** present. Being an interactive dashboard selecting any property class will display the data of that property on all the 5 tiles. For instance, selecting "private college, University " property will display the average assessed value as **$373.9** million and the total number of units as **2.9K** of the selection on the KPI tiles. Users can select more than one property at a time to observe the same analysis done for the selection. The two KPIs, the Line graph and Pie chart, will display the results for an aggregate of these selected multiple properties.

**Line Graph**

The line graph tile located in the center of the right side half of the dashboard exhibits the trend of the **assessed values from the year 2016 to 2021.** The X-axis represents the average assessment value ranging from **$6** million to **$12** million. The Y-axis represents each assessment year from **2016** to **2021**. The average assessment values of the top **10** categories started at **$7** million in the year **2016** and gradually increased to **$10** million until the year **2020** and then plateaued.

**Donut Chart**

The Donut chart located in the bottom right corner demonstrates the breakup of the top 10 properties of selected properties in percentages based on the count of the property category. The finding indicates that the majority of the properties around Cambridge are condominiums.

**Design Decision**

For the first visualization, we decided to use a **bar chart** to depict the assessed value of the **top 10** residential property categories because we wanted to illustrate how the assessed value is ranked in each residential type. In addition, to make it simpler for the audience to see the difference in average assessed value between property classes, we oriented the bar graph horizontally so that the audience could distinguish between the proportions in each property class. For instance, the visual indicates that the **MULTIUSE-IND** is approximately one-third larger than the **HIGH-TECH**. Furthermore, to lower the cognitive load of the chart, we reduced the grid line and expanded the Y-axis scale since the label already conveys the data. The second graph represents residential **assessed value** by year. We utilized a line graph to illustrate the trend of the assessed value throughout the year. Consequently, the audience will clearly see that the trend increases consistently throughout the year. Then, we illustrate the proportion of the property class categories using a **donut chart** to better comprehend the percentage of the total number of units in each category. This visual allows the audience to quickly determine which property type has the biggest volume in the Cambridge region. As the property class is already displayed on the label, we also omit the legend from the chart's look to lessen the cognitive load. Lastly, we designed a **KPI chart** to illustrate the average residential assessed value and the total number of residents to help the audience appreciate the Cambridge residential market as a whole.

**CONCLUSION**

From dashboard 1, we can conclude that C2, C1, and C3 Tax Districts have the highest tax amount generated. We have used a bar graph, pie chart, and 2 KPIs to visualize our analysis. We can also infer that there are only commercial properties in the C2 tax district. C5 Tax District has 99% of commercial properties whereas 1% is residential properties. This shows that **C2 and C5** districts are more confined to **commercial** properties than residential properties. **R7, R3, and R6** districts have more residential properties compared to commercial with a percentage over 60%. Here, we can conclude that these districts are **ideal for a residential spot**. From dashboard 2, we can say that the Assessed value for Private colleges is the highest. This can be evident because Massachusetts is considered an educational hub and assessment for university properties will be the highest. The **assessment value has increased over the years** (2016-2021) for all the property types. Out of all the property types in Cambridge, **42.1%** of the properties are **condominiums** and **12.8%** of the properties are **single-family residential buildings.**

**REFERENCES**

*Cambridge Property Database FY2016-FY2022: Open Data: City of Cambridge.* (2021, December 1). Cambridge Open Data. Retrieved from: <https://data.cambridgema.gov/Assessing/Cambridge-Property-Database-FY2016-FY2022/eey2-rv59>

Curt, Frye. (2019, March 7). *Qlik Sense Essential Training*. DataFlair: <https://www.linkedin.com/learning/qlik-sense-essential-training/join-data-source-with-different-field-names?autoplay=true&u=74653650>