**Data Analysis for Property Tax in DAP(City of Vancouver)**

Using AWS and data analytics tools, the Property Tax Project analyzes and visualizes property tax data from 2023 and 2024 to find trends, implications, and practical insights for stakeholders. This project seeks to improve knowledge of tax dynamics across several property kinds and areas, thereby provide data-driven recommendations for next decisions.

**Step 1: Formulate Focused Data Analytical Questions**

The next step was to formulate focused analytical questions to guide the analysis with respect to both 2023 and 2024 Property Tax datasets as follows:

* **Year-over-Year Trends:** How are property tax values changing from 2023 to 2024 across different property types and zoning classifications?
* **Zoning Impact:** What was the impact of zoning classifications on property tax assessments in 2023 and 2024?
* **Regional Differences:** Are the assessed property tax values varied across regions between these two years?
* **Comparison of Property Types:** How are tax values across different property types, like strata and land, changing over these years?

These questions provide a basis for understanding some key aspects of the property tax data and give insights for stakeholders that become very useful in guiding further data processing and analysis.

A screenshot of a computer

Description automatically generated

*Fig 1: Analysing structured Data*

**Step 2: Data Discovery**

The 2023 and 2024 Property Tax datasets were downloaded from the City of Vancouver Open Data Portal in Excel format.

Key Tasks are:

* **Dataset Overview:** The structure was understood with a view toward fields such as property type, zoning classification, and tax values.
* **Data Profiling:** To understand the pattern of the data in order to find anomalies. Descriptive
* **Check for missing/incomplete data:** Check for missing or incomplete data.
* **Early Insights:** Preliminary insights about first impressions.

This phase provided a general understanding of the datasets and hence laid the basis for effective analysis.

A screenshot of a computer

Description automatically generated

*Fig 1: Data Source and collecting from City of Vancouver Dashboard*

**Step 3: Design of Data Storage**

Designed a storage solution using AWS to manage the datasets as follows:

* **Storage Solution:** Excel files stored in Amazon S3 for scale and easy access.
* **Data Organizing:** Datasets were organized into different S3 buckets for the years 2023 and 2024 inside the landing zones.
* **Security:** Applied access controls and encryption.
* **Accessibility:** Gave access to analysis via AWS Glue and Amazon Athena. This ensured that data was in a proper structured format, secure, and ready to be processed.

A screenshot of a computer

Description automatically generated

*Fig 1: Data organising in landing zone of S3 bucket in AWS in excel format*

**Step 4: Dataset Preparation**

The datasets used for analysis were prepared with the aid of AWS Glue DataBrew as follows:

* **Data Cleaning:** Handling missing values, inconsistency correction (Descriptive).
* **Data Transformation:** Added a year column for the facilitation of comparison.
* **Data Storage:** Clean datasets stored in their corresponding Raw S3 bucket folders.

This step ensured the cleaning, organization, and preparation of the data for analysis.

A computer screen shot of a computer screen

Description automatically generated

*Fig 1: Data handling in AWS Glue Draw for 2023 dataset*

A screenshot of a computer

Description automatically generated

*Fig 1: Data handling in AWS Glue Draw for 2024 dataset*

A screenshot of a computer

Description automatically generated

*Fig 1: Cleaned Data(2023,2024) stored in Raw Data folder in their respective locations in CSV*

**Step 6: Design a Data Pipeline**

Design an ETL pipeline in Draw.io that can process data transformation and loading:

* **Flow Plan**: Mapping of the data flow from the Landing S3 to the Curated S3 bucket.
* **Design Transformation:** Describe the Transformations to be done on the data, such as the addition of year columns, and aggregate tax levy dataset.
* **Tool Selection:** AWS Glue for ETL.
* **Visual Workflow:** Draw.io draws a visual workflow.

This phase of design delivered a well-structured pipeline, ready for implementation.

A screenshot of a computer

Description automatically generated

*Fig 1: ETL Pipeline Design for Property Tax report using Draw.io*

**Step 7: Implement Data Pipeline using AWS Glue**

* **Extracting Data:** Extracted data from the Raw S3 bucket.
* **Filtering Columns:** Removing columns not needed, only relevant data is retained.
* **Schema Changes:** Renaming columns for consistency.
* **Merging Data:** Combination of datasets to have a unified analysis—a descriptive one.
* **Aggregation:** Calculated the average tax levy per year.
* **Union of Data:** Combined datasets using union.
* **Data Ingestion:** Loaded transformed data into the Curated S3 bucket.

This step ensured that the data was transformed, merged, and organised according to the object of the project.

A screenshot of a computer

Description automatically generated

*Fig 1: Implementing designed ETL pipeline in AWS Glue*

A screenshot of a computer

Description automatically generated

*Fig 1: Storing Implemented Data in Curated zone of S3 bucket in CSV form*

**Step 8: Analysing Data**

Used Amazon Athena in the analysis of the transformed datasets.

* **Table Creation:** Attached External tables in Athena to the curated S3 data.
* **Running Queries:** Using SQL to run the queries
* **Descriptive:** Trends of aggregated data.
* **Diagnostic:** Causes of trends observed were investigated
* **Predictive:** Property tax trends were forecasted to the future
* **Validation:** Accuracy of query results checked
* **Exporting Results**: Important results exported for visualization.

This step effectively gave valuable insights using the querying power of Athena.

A screenshot of a computer

Description automatically generated

*Fig 1: Dataset preparation using AWS Anthena*

A screenshot of a computer

Description automatically generated

*Fig 1: Data Table preparation using SQL query in AWS Anthena*

**A screenshot of a computer

Description automatically generated**

*Fig 1: Soring Database in Dataproducts of S3 bucket*

**Step 9: Data Visualization**

Due to a limited budget, Excel was utilized for this step in data visualization.

* **Selection of Tool:** Excel, due to its capability of charting.
* **Generation of Charts**: Pie, bar, and line charts were generated to visualize insights.
* **Dashboard Designing:** All the above-mentioned charts were placed in an Excel dashboard in order to make them interactive.
* **Presentation:** That dashboard was prepared in the form of a PDF report for stakeholders.

How this approach clearly and actionably presented results of analysis.

A screenshot of a graph

Description automatically generated

*Fig 1: Preparing Data Charts in EXCEL Dashboard*

**Step 10: Data Publishing**

Setting up Server Instances for Storing and Publishing Results:

* **General Server Instance:** It held the processed data and visualizations.
* **Web Server Instance:** Findings were hosted via a web interface.
* **Data Storage:** Upload of datasets and reports for accessing from one location.
* **Web Publishing:** The report is accessible via a browser interface.
* **Prescriptive Analytics:** Recommendations on the basis of analysis for decision-making.

This step insured that the findings are stored securely, published efficiently, and easily accessed by the stakeholders.

A screenshot of a computer

Description automatically generated

*Fig 1: Creating EC2 Instances in AWS cloud*

A screenshot of a computer

Description automatically generated

*Fig 1: Storing Database in Back-End Infrastructure*

A screenshot of a computer

Description automatically generated

*Fig 1: Visualizing Data Base in back-end*

A screenshot of a computer

Description automatically generated

*Fig 1: Auto-assigned Public IP address of Web Server(Front end)*

A screenshot of a computer

Description automatically generated

*Fig 1: Visualizing Data Base in remote access(front-end for the client)*