# Data-Analyst-Omar

# Table of Contents

1. Exploratory Data Analysis

- Project Description

- Project Title

- Objective

- Dataset

- Methodology

- Tools and Technologies

- Deliverables

2. Descriptive Analysis

- Project Description

- Project Title

- Objective

- Dataset

- Methodology

- Tools and Technologies

- Deliverables

3. Diagnostic Analysis

- Project Description

- Project Title

- Objective

- Background

- Dataset

- Methodology

- Tools and Technologies

- Deliverables

- Timeline

4. Data Wrangling

- Project Description

- Project Title

- Objective

- Background

- Dataset

- Methodology

- Tools and Technologies

- Deliverables

- Timeline

5. Data Quality Control

- Project Description

- Project Title

- Objective

- Background

- Scope

- Methodology

- Deliverables

- Timeline

6. Course Completion Badge

# Exploratory Data Analysis

Project Description

An analysis of the property tax dataset aims to find patterns together with inconsistencies and abnormal cases for improving municipal assessment methods.

Project Title

Exploratory Data Analysis of Property Tax Data

Objective

The main purpose of this project involves analyzing the property tax dataset while determining tax value patterns and searching for irregularities within tax evaluation procedures to offer evidence-based guidance to policymakers. The analysis focuses on evaluating different property aspects with their tied tax assessments to maximize transparency and decision quality within municipal tax frameworks.

Dataset

The property tax database includes extensive information about tax assessments together with property-related data. It includes the following attributes:

* Property ID: Unique identifier for each property
* Location: Geographic location of the property
* Property Type: Classification of the property (e.g., residential, commercial, industrial)
* For taxation purposes the property obtains its valuation through estimated market worth calculations.
* The total amount of property tax that arises from assessment calculations represents the tax amount.
* Assessment Year demonstrates the specific period when local authorities performed the assessment process.
* Property ownership status shows whether occupants manage the property or a tenant occupies it.
* The historical tax data contains previous property tax amounts alongside assessment change records which forms the base for fair tax assessment analysis and anomaly discovery.

Methodology

1- Data Cleaning and Preprocessing:

* The first step involves loading property tax data followed by primary assessment of information in this dataset.
* The process handles data problems in crucial tax payment field entries including assessed values and tax amounts.
* The dataset requires standardized property classification systems with uniform data field conventions.

2- Statistical Analysis and Data Distribution:

* The data analysis includes calculating the average value through three measures such as mean value and median value and standard deviation of property assessments and taxes.
* Data scientists will examine the distribution patterns of data that includes properties grouped by type and settled in different locations.
* The analysis detects tax assessment outliers through statistical methods that use interquartile range (IQR) and z-scores.

3- Identifying Trends and Anomalies:

* We should employ data visualization methods for visualizing tax value shifts from year to year as well as between different property categories.
* The assessment inconsistency detection process includes the comparison of assessments between identical properties.
* The analysis should present how policy implications appear through observed differences in property tax distribution.

Tools and Technologies

- data processing, statistical analysis, and visualization.

- Tableau for Visualization

Deliverables

* Data Summary Report
* Trend and Outlier Analysis
* Visualized Insights

# Descriptive Analysis

Project Description

This research studies how property tax values distribute through evaluations of different property types and location effects and valuations changes. The research finds tax assessment patterns and irregularities which enable officials to create valuable knowledge for property owners and analysis professionals as well as government authorities.

Project Title

Descriptive Analysis of Property Tax Data

Objective

The main goal of this investigation entails the assessment of key property tax value statistics while investigating variations that stem from various properties across multiple locations. Tax assessments receive evaluation support because the analysis shows both patterns and inconsistent data which helps determine their accuracy level.

Dataset

The available records include tax data for properties as follows:

* Property ID – Unique identifier for each property
* Location – Geographic area of the property
* Property Type – Classification such as residential, commercial, or industrial
* For tax computation purposes the assessment personnel uses the market-valued properties to determine their value.
* The levied tax equals the calculated assessment value.
* Property owners must use Assessment Year indicators to determine the time when tax assessments occurred.
* Historical Tax Data – Changes in tax amounts over multiple years

Methodology

1.Data Aggregation and Summarization

* The dataset requires cleaning up and preprocessing to resolve data entry issues with missing or inconsistent information.
* The summary statistics for both tax amounts and assessed values consist of mean averages and median values together with standard deviations.

2.Comparative Analysis by Location and Property Type

* The analysis examines the differences between tax amounts which occur due to the varying geographic locations.
* An analysis should be conducted to detect if property tax distributions between types show significant variations.

3.Statistical Visualization

* This system uses Histograms and Boxplots to show the distribution pattern and measurement spans of tax values.
* Heatmaps: Identify geographic disparities in property tax assessments.
* The bar chart analysis shows the amount of tax variations between various property categories.

Tools and Technologies

* data processing, statistical analysis, and visualization.
* Tableau: Employed for interactive dashboards and graphical representations of tax trends.

Deliverables

* The Summary Statistics Report presents the main outcomes regarding tax distribution patterns.
* The platform provides an overview of tax variations that detect differences between various property types and geographical areas.

Diagnostic Analysis

Project Description

Examinations of non-alignment between property tax appraisals need to connect with elements which lead to variations in tax amounts.

Project Title

The research investigates the origins of discrepancies found in property taxation.

Objective

This project dedicates its core purpose to running a diagnostic investigation which will reveal fundamental sources behind tax assessment irregularities. The tax record analysis seeks to determine which factors cause discrepancies so tax authorities can develop effective methods to improve their assessment accuracy.

Background

Taxpayers across the previous year identified various inconsistent properties on their tax assessment records. Property tax valuation discrepancies between different owners have triggered worries which require a deep evaluation of the factors that cause such value discrepancies. By performing the analysis the study will both discover assessment process origins and present procedures to enhance evaluation assessments.

Dataset

This analysis depends on three different datasets as part of its research process.

* The dataset features Property Tax Data which includes property assessment values with property characteristics such as dimensions and location and tax levies collected from the City of vancouver.
* The dataset contains Property Feature Data that describes essential characteristics of properties including their dimensions, use types, construction type and age of the property.
* The evaluation includes market data which combines economic measurements like property dealing prices along with neighborhood-trend patterns combined with area statistics for determining property market worth.
* Assessment Discrepancy Reports: Complaints and feedback from property owners about discrepancies in their tax assessments.

Methodology

1.Data Collection and Preparation:

* Declutter and normalize all datasets obtained from different sources to establish uniformity along with factual precision.
* The data receives normalization treatment to achieve relative equality between properties of different types and areas.

2.Outlier Detection:

* A statistical analysis involving Z-scores and IQR analysis should be used for outlier detection in property tax assessments to locate irregular tax values which could point to assessment discrepancies.

3.Correlation Analysis:

* The relationship analysis will establish significant factors which affect tax values based on characteristics like location or property size to determine assessment discrepancies.
* This analysis uses regression to establish the degree of influence which influencing factors have on assessment values.

4.Data Quality Investigation:

* Research any issues which exist in data elements including empty spaces or incorrect entries because they might generate assessment value misunderstandings.

5.Segmentation Analysis:

* The comparison of property assessment discrepancies should be conducted through categorization of properties into residential and commercial sections to detect variations between different groups.

6.Root Cause Analysis:

* Tax assessors and local officials need to be interviewed to obtain information about the assessment procedures and potential assessment errors and biases.

7.Synthesis of Findings:

* A synthesis of quantitative data with qualitative research needs to be used to identify the primary elements behind property tax irregularities.

Tools and Technologies

* The Python programming language with Pandas and Matplotlib along with Scikit-learn as libraries functions to process data which conducts analysis operations.
* One of the tools used for finding presentation and stakeholder trend visualization was Tableau.

Deliverables

* The diagnostic report summarizes the analytical approach and reveals assessment inconsistency causes together with all identified findings.
* The report includes section visualizations and dashboard displays of essential data patterns alongside main elements from the analysis findings.
* Actions steps will be provided regarding how tax authorities should enhance property tax assessment accuracy and consistency.

Timeline

* Data Collection & Preparation: Feb 28 – Feb 30
* Outlier Detection & Analysis: Feb 31 - Feb 2
* Correlation and Segmentation Analysis: Mar 3 - Mar 5
* Root Cause Analysis & Synthesis: Mar 6 - Mar 8
* Final Report & Deliverables: Mar 9 - Mar 11

# Data Wrangling

Project Description

A cleaning process and structural normalization of the property tax dataset from Vancouver was performed to enhance analysis capability and prevent tax discrepancies in the system.

Project Title

Data Wrangling for Property Tax Dataset in Vancouver

Objective

* Data effects for analysis require proper consistency and accurate information along with complete resolution in place.
* We should remove all duplicate entries together with erroneous and missing data points.
* The classification system for properties needs standardization to simplify the comparison process between different types of properties across different regions.

Background

The raw property tax data in Vancouver poses multiple formatting problems which include gaps in values and duplicate entries together with specification inaccuracies. Proper data wrangling becomes necessary to solve these data quality problems before researchers can use the data for analysis of tax assessment discrepancies.

Dataset

The Vancouver property tax dataset contains information about tax records together with ownership data and property class types and assessments.

Methodology

1.Handling Missing and Incorrect Data:

* The approach will determine whether to fill in missing values through imputation or eliminate them based on the affected data's magnitude.
* The system needs to recognize and solve mistakes within its database including instances where values are negative or tax assessments fall outside normal parameters.

2.Standardizing Property Classification:

* A consistent method should be implemented to classify properties between residential and commercial and industrial types so the dataset remains uniform.
* All discrepancies in property types must be checked for accuracy before they match tax assessment classifications.

3.Data Formatting and Structuring:

* The team will modify unorganized data through proper formatting by adopting methods like date standardization and address normalization.
* The data needs proper organization for analysis through renaming columns and applying uniform measurement units to property size properties and tax values.

Tools and Technologies

* data cleaning and transformation
* AWS Glue serves as the data platform for all tasks related to preprocessing and cleaning and standardization of data.

Deliverables

* A cleaned version of the Property Tax Dataset became ready for use after processing.
* The report includes detailed descriptions of applied preprocessing methods together with a list of managed processing issues.

Timeline

* Data Cleaning: 1 week
* Transformation & Standardization: 2 weeks

# Data Quality Control

Project Description

Establishing robust data quality control measures to ensure the accuracy, completeness, and reliability of the property tax dataset from Vancouver. This project will focus on implementing quality checks to address data inconsistencies, improve decision-making, and ensure the integrity of property tax assessments.

Project Title

Implementation of Data Quality Control Measures for Vancouver Property Tax Dataset

Objective

The primary objective of this project is to establish a comprehensive Data Quality Control (DQC) framework for Vancouver’s property tax dataset. This framework will ensure the dataset's accuracy, completeness, consistency, and reliability, thereby enhancing the quality of property tax assessments and supporting accurate analysis for future decision-making.

Background

Vancouver’s property tax dataset contains errors, such as missing values, duplicate records, and inconsistent formatting. These data quality issues pose a risk to the reliability of tax assessments, potentially leading to financial discrepancies, disputes, and regulatory issues. As Vancouver’s property tax data continues to grow, the need for effective data quality control measures has become crucial to maintaining accurate assessments and ensuring the smooth operation of the tax system.

Scope

The project will focus on the following key areas:

* Data Profiling: Analyzing the existing property tax datasets to assess quality levels.
* Data Cleansing: Developing processes to correct inaccuracies, remove duplicates, and standardize property classification data.
* Data Validation: Implementing validation rules to ensure data integrity at the point of entry and throughout the dataset.
* Monitoring and Reporting: Establishing ongoing monitoring processes and dashboards to track key data quality metrics.
* Training and Awareness: Educating staff on data quality best practices for continuous improvement.

Methodology

1.Current State Assessment:

* An assessment should take place to review the current property tax database while evaluating both present data flaws and their basic origins together with missing or incorrect data instances.
* The assessment will focus on determining important data elements that affect tax assessment quality including tax values together with property types and ownership data points.

2.Data Profiling:

* The use of data profiling tools allows assessment of data quality through identification of missing data completeness and duplicates together with data validity validation and value consistency measurements and entry accuracy checks.
* Create a documented report of the observed problems to determine urgent maintenance needs followed by procedures for cleaning order.

3.Establish Data Quality Metrics:

* Errors rates along with duplicates and missing entries and tax standards adherence and intracompany data standards compliance will serve as metrics to assess and monitor the dataset’s quality.
* Key Performance Indicators (KPIs) should be deployed to track data reliability and accuracy development because they will monitor enhancements over time.

4.Data Cleansing Processes:Procedures must be implemented to fix data errors through the following actions:

* The database system implements a process to eliminate duplicate records through unique identifier matching like property ID.
* The procedure involves unifying property tax categories including residential and commercial zones as well as creating uniform property classification values.
* The system utilizes acceptable methods for filling empty data points with methods including mean imputation and value propagation for linked fields.
* Users must perform data correction activities involving tax value ranges and date formats as well as inconsistent tax assessment values.

5.Validation Rules and Procedures:

* The entire system will have validation rules to screen new property tax entries which must comply with established requirements related to format structure and data precision.
* The system should automatically detect new errors which occur during data entry activities to prevent poor-quality information from contaminating the database.

6.Monitoring and Reporting:

* The use of real-time data quality monitoring dashboards combined with tools should enable the organization to visualize important metrics track data quality trends during specific periods.
* The system should trigger notifications when quality standards encounter substantial deviations such as excessive missed values or raised error numbers.
* Staff should receive scheduled reports about data quality performance trends while observing key metrics that include tax assessment accuracy and property classification consistency alongside missing data.

7.Training and Best Practices:

* The organization should create educational materials for training staff members including tax assessors and data analysts about data quality standards and proper data entry protocols.
* Staff members should receive ongoing support to develop awareness about quality data through a system that promotes immediate identification and resolution of existing issues.
* The system requires a user-driven feedback procedure which will enhance data quality methods through employee observations and dataset performance measurements.

8.Feedback Mechanism:

* Staff members and data users should have an ongoing system to alert the team about quality problems and recommend improvements.
* The system development method includes periodic reviews to evaluate and optimize data quality procedures which allow the system to adjust to new operational requirements based on stakeholder feedback.

Tools and Technologies

* Data Quality Tools: Python (Pandas, NumPy, PyTest for validation scripts), AWS Glue, AWS DataBrew, or Talend for profiling, cleansing, and data validation.
* Data Visualization Tools: Tableau or Power BI for real-time monitoring and reporting on data quality metrics.
* Database Management: SQL for querying and updating the dataset based on the quality control measures.

Deliverables

* A comprehensive Data Quality Control Plan detailing the processes, metrics, and responsibilities involved in maintaining data quality.
* Documentation of the key data quality metrics (KPIs) being tracked, such as error rates, duplicates, and missing data.
* Cleaned and validated property tax dataset, ready for use in further analysis and reporting.
* Training materials and workshops designed to educate staff on data quality best practices.
* A monitoring dashboard that provides real-time visualization of data quality metrics, trends, and issues.

Timeline

* Current State Assessment and Data Profiling: 2 weeks
* Data Cleansing and Validation Setup: 3 weeks
* Monitoring and Reporting Setup: 2 weeks
* Training and Best Practices: 1 week
* Ongoing Monitoring and Refinement: Ongoing after the initial 1-month period

# Course Completion Badge

The graduate has finished coursework that focuses on data analysis combined with visualization techniques and data quality administration. The document displays ability in analyzing property tax data with specialized focus on quality control and data visualization.

<https://www.credly.com/earner/earned/share/acb682b0-513a-4806-84ea-b99912f95589>