People's Democratic Republic of Algeria

Ministry of Higher Education and Scientific Research

Saad Dahleb Blida 01 University

Department of computer science



MASTER’S INTELLIGENT SYSTEMS ENGINEERING

DATA ANALYSIS

|  |  |  |
| --- | --- | --- |
| STUDENTS : | |  |
| 01. | ABDELATIF MEKRI |
| 02. | HALIMA NFIDSA |
| 03. | NAHLA YASMINE MIHOUBI |
|  | | SUPERVISED BY : |
| Dr. M.FAREH |

1. Abstract

The primary objective of this mini-project is to conduct a comprehensive data analysis on a specified dataset. The analysis aims to explore, identify, and understand the relationships, patterns, and trends within the dataset. Through sequential and progressive analysis, the project seeks to extract valuable insights into the studied variables, their interactions, and mutual influences, thereby providing a thorough understanding of the data to facilitate informed decision-making.

To comprehend and interpret the data effectively, the project follows a systematic approach, encompassing univariate, bivariate, and multivariate analysis steps. These steps enable the exploration and interpretation of different facets of the data, starting from simpler analyses and progressing towards more complex and detailed examinations. Univariate, bivariate, and multivariate analyses are integral components, ensuring a methodical and in-depth execution of the data analysis project.

The project requires the utilization of the Python programming language to analyze a specific dataset. The student is tasked with implementing all three categories of analyses (univariate, bivariate, and multivariate) and subsequently compiling a detailed report outlining the adopted methodology for these analyses.

1. Introduction

Data analysis is a crucial aspect of deriving meaningful insights from complex datasets. In this project, the focus is on employing Python for a meticulous examination of a designated dataset. The analysis encompasses univariate, bivariate, and multivariate approaches, each contributing to a holistic understanding of the dataset's characteristics.

The initial step involves obtaining a dataset, commonly referred to as a dataset or dataset, which serves as the foundation for all subsequent analyses. This dataset comprises various observations and their corresponding attributes or variables, ranging from numerical (continuous or discrete) to categorical types. The project's scope involves describing the selected dataset, including details such as download links, the number of attributes, the number of individuals, and the types of variables, thereby setting the groundwork for the ensuing analyses.

Additionally, the project outlines three main tasks: stating the objectives to be achieved on the dataset, emphasizing the importance of preprocessing the dataset before analysis, and delineating the steps involved in preprocessing. The preprocessing phase is pivotal, involving tasks such as managing missing values, deduplication, data normalization, and transformations, ensuring that the dataset is coherent and prepared for in-depth analysis. Through this introduction, the project establishes a framework for a systematic and insightful exploration of the dataset, with a clear roadmap for subsequent analytical endeavors.

1. Objectives of the Project:
   1. Understanding House Price Trends:

Explore and analyze the historical trends in the FHFA House Price Index (FHFA HPI) to gain insights into the movement of single-family house prices over time.

Identify periods of significant growth or decline in house prices at the national and regional levels.

* 1. Geographic Analysis:

Conduct a comprehensive geographic analysis by examining house price fluctuations at different levels, including census divisions, states, metro areas, counties, ZIP codes, and census tracts.

Identify regions with notable variations in house price trends and understand the factors contributing to these variations.

* 1. Methodology Understanding:

Gain a deep understanding of the HPI methodology, which utilizes a weighted, repeat-sales statistical technique.

Explore how the HPI incorporates data from tens of millions of home sales, providing a transparent and comprehensive approach to analyzing house price transaction data.

* 1. Economic Indicators:

Assess the HPI as an economic indicator, understanding its relevance in estimating changes in mortgage defaults, prepayments, and housing affordability.

Investigate how the HPI can be utilized as a tool for housing economists to enhance analytical capabilities in specific geographic areas.

* 1. Dataset Exploration and Description:

Provide a detailed description of the chosen dataset, including the download link, the number of attributes, the number of observations (house price transactions), and the types of variables (e.g., geographic, temporal).

Understand the structure and composition of the HPI dataset to ensure effective analysis.

1. Reasons for Choosing the HPI Dataset:

Comprehensive Coverage:

The FHFA HPI dataset offers a comprehensive collection of publicly available house price indexes, covering all 50 states and over 400 American cities.

The extensive coverage provides a rich source of data for a thorough analysis of house price trends at various geographic levels.

Longitudinal Data:

The dataset extends back to the mid-1970s, providing a longitudinal perspective on house price changes.

Longitudinal data enables the identification of historical patterns and facilitates a more robust analysis of the factors influencing house prices over time.

Transparency and Methodology:

The FHFA HPI employs a fully transparent methodology based on a weighted, repeat-sales statistical technique.

The transparent methodology enhances the credibility of the dataset, ensuring a reliable basis for analysis and interpretation.

Insights for Decision-Making:

The FHFA HPI serves as a timely and accurate indicator of house price trends, offering valuable insights for decision-making in real estate, finance, and economic sectors.

The dataset's ability to provide information at various geographic levels enhances its utility for localized decision-making.

Relevance to Housing Economists:

The FHFA HPI dataset is relevant to housing economists, providing an improved analytical tool for estimating changes in mortgage defaults, prepayments, and housing affordability.

The dataset's granularity allows for nuanced analyses, contributing to a better understanding of economic trends in specific regions.

By pursuing these objectives and leveraging the strengths of the FHFA HPI dataset, the project aims to contribute valuable insights into the dynamics of house prices and their economic implications.