People's Democratic Republic of Algeria

Ministry of Higher Education and Scientific Research

Saad Dahleb Blida 01 University

Department of computer science



MASTER EN INGÉNIERIE DES SYSTÈMES INTELLIGENTES

ANALYSE DE DONNÉES

|  |  |  |
| --- | --- | --- |
| ÉTUDIENTS : | |  |
| 01. | ABDELATIF MEKRI |
| 02. | HALIMA NFIDSA |
| 03. | NAHLA YASMINE MIHOUBI |
|  | | SUPERVISÉ PAR : |
| Dr. M.FAREH |

1. Résumé

L'objectif principal de ce mini-projet est de mener une analyse complète des données sur un ensemble de données spécifié. L'analyse vise à explorer, identifier et comprendre les relations, les modèles et les tendances au sein de l'ensemble de données. À travers une analyse séquentielle et progressive, le projet cherche à extraire des insights précieux sur les variables étudiées, leurs interactions et leurs influences mutuelles, fournissant ainsi une compréhension approfondie des données pour faciliter la prise de décision éclairée. Pour comprendre et interpréter les données de manière efficace, le projet suit une approche systématique, englobant des étapes d'analyse univariée, bivariée et multivariée. Ces étapes permettent l'exploration et l'interprétation de différentes facettes des données, commençant par des analyses plus simples et progressant vers des examens plus complexes et détaillés. Les analyses univariées, bivariées et multivariées sont des composants intégraux, garantissant une exécution méthodique et approfondie du projet d'analyse de données. Le projet nécessite l'utilisation du langage de programmation Python pour analyser un ensemble de données spécifique. L'étudiant est chargé de mettre en œuvre les trois catégories d'analyses (univariée, bivariée et multivariée) et de compiler ensuite un rapport détaillé décrivant la méthodologie adoptée pour ces analyses.

1. Introduction

L'analyse de données est un aspect crucial pour tirer des insights significatifs de jeux de données complexes. Dans ce projet, l'accent est mis sur l'utilisation de Python pour un examen méticuleux d'un ensemble de données désigné. L'analyse englobe des approches univariées, bivariées et multivariées, contribuant chacune à une compréhension holistique des caractéristiques de l'ensemble de données. La première étape implique l'obtention d'un ensemble de données, communément appelé jeu de données, qui sert de base à toutes les analyses ultérieures. Cet ensemble de données comprend diverses observations et leurs attributs ou variables correspondants, allant des types numériques (continus ou discrets) aux types catégoriels. La portée du projet consiste à décrire l'ensemble de données sélectionné, y compris des détails tels que les liens de téléchargement, le nombre d'attributs, le nombre d'individus et les types de variables, établissant ainsi les bases pour les analyses ultérieures. De plus, le projet détaille trois tâches principales: énoncer les objectifs à atteindre sur l'ensemble de données, souligner l'importance du prétraitement de l'ensemble de données avant l'analyse et détailler les étapes du prétraitement. La phase de prétraitement est cruciale, impliquant des tâches telles que la gestion des valeurs manquantes, la déduplication, la normalisation des données et les transformations, garantissant que l'ensemble de données soit cohérent et prêt pour une analyse approfondie. À travers cette introduction, le projet établit un cadre pour une exploration systématique et perspicace de l'ensemble de données, avec une feuille de route claire pour les efforts analytiques ultérieurs.

1. Objectifs du Projet:
   1. Compréhension des Tendances des Prix des Maisons: Explorer et analyser les tendances historiques de l'Indice des Prix des Maisons (HPI) pour obtenir des insights sur l'évolution des prix des maisons individuelles au fil du temps. Identifier les périodes de croissance ou de déclin significatives des prix des maisons aux niveaux national et régional.
   2. Analyse Géographique: Mener une analyse géographique complète en examinant les fluctuations des prix des maisons à différents niveaux, y compris les divisions de recensement, les états, les zones métropolitaines, les comtés, les codes ZIP et les secteurs de recensement. Identifier les régions avec des variations notables dans les tendances des prix des maisons et comprendre les facteurs contribuant à ces variations.
   3. Compréhension de la Méthodologie: Acquérir une compréhension approfondie de la méthodologie du HPI, qui utilise une technique statistique pondérée de ventes répétées. Explorer comment le HPI intègre les données de dizaines de millions de ventes de maisons, fournissant une approche transparente et complète pour analyser les données de transactions de prix des maisons.
   4. Indicateurs Économiques: Évaluer le HPI en tant qu'indicateur économique, en comprenant sa pertinence pour estimer les changements dans les défauts de prêts hypothécaires, les remboursements anticipés et l'accessibilité au logement. Enquêter sur la manière dont le HPI peut être utilisé comme un outil pour les économistes du logement afin d'améliorer les capacités analytiques dans des zones géographiques spécifiques.
   5. Exploration et Description de l'Ensemble de Données: Fournir une description détaillée de l'ensemble de données choisi, y compris le lien de téléchargement, le nombre d'attributs, le nombre d'observations (transactions de prix des maisons) et les types de variables (par exemple, géographiques, temporelles). Comprendre la structure et la composition de l'ensemble de données HPI pour assurer une analyse efficace.
2. Raisons du Choix de l'Ensemble de Données HPI:

Couverture Complète: L'ensemble de données HPI de la FHFA offre une collection complète d'indices des prix des maisons disponibles publiquement, couvrant les 50 États et plus de 400 villes américaines. La couverture étendue fournit une source riche de données pour une analyse approfondie des tendances des prix des maisons à différents niveaux géographiques.

Données Longitudinales: L'ensemble de données remonte au milieu des années 1970, offrant une perspective longitudinale sur les changements de prix des maisons. Les données longitudinales permettent l'identification de motifs historiques et facilitent une analyse plus robuste des facteurs influençant les prix des maisons au fil du temps.

Transparence et Méthodologie: Le HPI de la FHFA utilise une méthodologie entièrement transparente basée sur une technique statistique pondérée de ventes répétées. La méthodologie transparente renforce la crédibilité de l'ensemble de données, assurant une base fiable pour l'analyse et l'interprétation.

Perspectives pour la Prise de Décision: Le HPI de la FHFA sert d'indicateur opportun et précis des tendances des prix des maisons, offrant des insights précieux pour la prise de décision dans les secteurs de l'immobilier, des finances et de l'économie. La capacité de l'ensemble de données à fournir des informations à différents niveaux géographiques renforce son utilité pour la prise de décision localisée.

Pertinence pour les Économistes du Logement: L'ensemble de données HPI de la FHFA est pertinent pour les économistes du logement, fournissant un outil analytique amélioré pour estimer les changements dans les défauts de prêts hypothécaires, les remboursements anticipés et l'accessibilité au logement. La granularité de l'ensemble de données permet des analyses nuancées, contribuant à une meilleure compréhension des tendances économiques dans des régions spécifiques. En poursuivant ces objectifs et en exploitant les forces de l'ensemble de données HPI de la FHFA, le projet vise à apporter des insights précieux sur la dynamique des prix des maisons et leurs implications économiques.