Course End Project: Data Analysis Module

Exploratory Data Analysis

**Objective:**

The objective of this project is to conduct an exploratory data analysis (EDA) on a dataset of your choice. You have the option to use the Airbnb Listings & Reviews dataset or another dataset of your preference, subject to confirmation with the instructor. The goal remains the same: to analyze the dataset, understand its characteristics, and draw meaningful insights from it.

الهدف من هذا المشروع هو إجراء تحليل البيانات الاستكشافية (EDA)

على مجموعة بيانات من اختيارك. لديك خيار استخدام مجموعة بيانات

Airbnb Listings & Reviews أو مجموعة بيانات أخرى تفضلها، ويخضع ذلك للتأكيد مع المدرب. ويظل الهدف كما هو: تحليل مجموعة البيانات، وفهم خصائصها، واستخلاص رؤى ذات معنى منها

* **Dataset Options:** خيارات مجموعة البيانات:
* [Airbnb Listings & Reviews Dataset](https://www.kaggle.com/datasets/mysarahmadbhat/airbnb-listings-reviews/data): This dataset contains information about Airbnb listings, including property details, host information, pricing, availability, and guest reviews. It comprises data for 250,000+ listings in 10 major cities, along with over 5 million historical reviews.
* • مجموعة بيانات قوائم ومراجعات Airbnb: تحتوي مجموعة البيانات هذه على معلومات حول قوائم Airbnb، بما في ذلك تفاصيل الملكية، ومعلومات المضيف، والأسعار، والتوفر، وتقييمات الضيوف. وهو يشتمل على بيانات لأكثر من 250.000 قائمة في 10 مدن رئيسية، إلى جانب أكثر من 5 ملايين مراجعة تاريخية.
* Alternate Dataset: You may propose an alternative dataset of your choice, but you must confirm its suitability with the instructor. The dataset should be sufficiently rich in attributes and suitable for exploratory data analysis.
* • مجموعة البيانات البديلة: يمكنك اقتراح مجموعة بيانات بديلة من اختيارك، ولكن يجب عليك التأكد من ملاءمتها مع المدرب. يجب أن تكون مجموعة البيانات غنية بما فيه الكفاية بالسمات ومناسبة لتحليل البيانات الاستكشافية

**Tasks:**

### Data Loading and Initial Exploration:تحميل البيانات والاستكشاف الاولي

* Load the selected dataset into your preferred data analysis environment. • تحميل مجموعة البيانات المحددة في بيئة تحليل البيانات المفضلة لديك
* Explore the structure of the dataset (number of rows and columns, data types, etc.). • استكشاف بنية مجموعة البيانات (عدد الصفوف والأعمدة، وأنواع البيانات، وما إلى ذلك).
* Examine the first few rows of the dataset to understand its contents. • افحص الصفوف القليلة الأولى من مجموعة البيانات لفهم محتوياتها.

### 2. Data Cleaning:تنظيف البيانات

* Handle missing values appropriately (e.g., imputation, deletion, etc.). • التعامل مع القيم المفقودة بشكل مناسب (على سبيل المثال، الإسناد، الحذف، وما إلى ذلك).
* Check for and handle any duplicate entries. • التحقق من وجود أي إدخالات مكررة والتعامل معها
* Convert categorical variables into the appropriate data type if necessary. • تحويل المتغيرات الفئوية إلى نوع البيانات المناسب إذا لزم الأمر

### 3. Exploratory Data Analysis: تحليل البيانات الاستكشافيه

* Conduct univariate analysis to understand the distribution of individual variables (e.g., listing prices, number of reviews, etc.). • إجراء تحليل أحادي المتغير لفهم توزيع المتغيرات الفردية (على سبيل المثال، قائمة الأسعار، وعدد المراجعات، وما إلى ذلك).
* Perform bivariate analysis to explore relationships between variables (e.g., price vs. property type, price vs. neighborhood, etc.). • إجراء تحليل ثنائي المتغير لاستكشاف العلاقات بين المتغيرات (على سبيل المثال، السعر مقابل نوع العقار، السعر مقابل الحي، وما إلى ذلك).
* Visualize the data using appropriate plots (e.g., histograms, box plots, scatter plots, etc.). • تصور البيانات باستخدام المخططات المناسبة (على سبيل المثال، الرسوم البيانية، المخططات المربعة، المخططات المبعثرة، وما إلى ذلك).
* Calculate summary statistics for relevant variables. • حساب إحصائيات موجزة للمتغيرات ذات الصلة

### 4. Feature Engineering: هندسة الميزات

* Create new features that might be useful for analysis (e.g., average rating, occupancy rate, etc.). • إنشاء ميزات جديدة قد تكون مفيدة للتحليل (على سبيل المثال، متوسط التصنيف، ومعدل الإشغال، وما إلى ذلك).
* Encode categorical variables if necessary (e.g., one-hot encoding, label encoding, etc.). • تشفير المتغيرات الفئوية إذا لزم الأمر (على سبيل المثال، التشفير السريع، تشفير الملصقات، وما إلى ذلك).

### 5. Hypothesis Testing (Optional): .اختبار الفراضيات (اختياري )

* Formulate hypotheses related to the dataset based on different factors. • صياغة الفرضيات المتعلقة بمجموعة البيانات بناءً على عوامل مختلفة.
* Perform statistical tests (e.g., t-test, ANOVA, etc.) to test these hypotheses. • إجراء اختبارات إحصائية (على سبيل المثال، اختبار t، ANOVA، وما إلى ذلك) لاختبار هذه الفرضيات

### 6. Conclusion:

* Summarize the findings from the exploratory data analysis. • تلخيص النتائج من تحليل البيانات الاستكشافية
* Discuss any insights or patterns observed in the data. • مناقشة أي رؤى أو أنماط لوحظت في البيانات.
* Make recommendations based on the analysis conducted. • تقديم توصيات بناء على التحليل الذي تم إجراؤه.

**Deliverables:الإنجازات**

1. Jupyter Notebook (or equivalent) containing all the code and analysis 1. Jupyter Notebook (أو ما يعادله) يحتوي على جميع الأكواد والتحليلات.
2. Visualizations (plots, graphs) to support the analysis. . تصورات (المؤامرات والرسوم البيانية) لدعم التحليل.
3. A brief report summarizing the findings and conclusions. 3. تقرير مختصر يلخص النتائج والاستنتاجات

**Note:**

* If choosing the alternate dataset option, you must confirm its suitability with the instructor before proceeding.
* Make sure to document your code and analysis steps clearly for better understanding.
* You can refer to online resources or books for guidance on data analysis techniques and best practices.
* Collaboration with peers or seeking guidance from the instructor is encouraged during the project.
* • في حالة اختيار خيار مجموعة البيانات البديلة، يجب عليك التأكد من ملاءمتها مع المدرب قبل المتابعة.
* • تأكد من توثيق التعليمات البرمجية وخطوات التحليل بشكل واضح من أجل فهم أفضل.
* • يمكنك الرجوع إلى الموارد أو الكتب عبر الإنترنت للحصول على إرشادات حول تقنيات تحليل البيانات وأفضل الممارسات.

• يتم تشجيع التعاون مع أقرانهم أو طلب التوجيه من المدرب خلال المشروع.