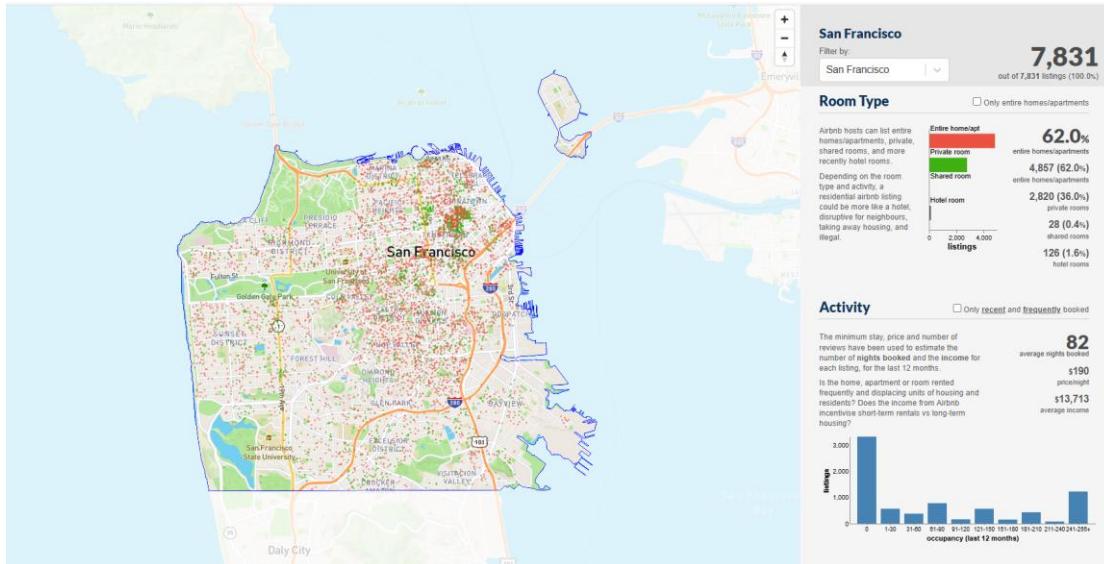


מיני פרויקט: יוז נדל"ן חכם בסן פרנסיסקו

שילוב של Machine Learning + Streamlit



 מטרה:

בנו מערכת שלמה שמקבלת פרטיים על דירה להשכלה בסן פרנסיסקו ומבצעת:

- חיזוי מחיר השכלה ללילה.
- הסבר בשפה פשוטה למה המחיר זהה.
- הציג המידע בדשبورד אינטראקטיבי (Streamlit).

את המילון שמספרנו את המשתנים ניתן למצוא כאן:

<https://docs.google.com/spreadsheets/d/1iWCNJcSutYqpULSQHINyGInUvHg2BoUGoNRIGa6SzC4/edit?gid=1322284596#gid=1322284596>

שלב 1: ניקוי נתונים (Data Cleaning)

מטרות:

- הסרת עמודות לא רלוונטיות.
- טיפול בערכים חסרים (NaN).
- המרה של עמודת מחיר מפורמט טקסט למספר.
- סינון קיצונים לא הגיוניים במחיר.

משימות לביצוע:

1. לטעון את קובץ הנתונים המקורי.
 2. לבחור רק את העמודות החשיבות למודל: מיקום, סוג חדר, מספר חדרים, אמצעי, מחיר וכו'!
 3. לבדוק האם בעמודת המחיר יש סימן \$ ולהסיר אותו.
 4. להמיר את המחיר לערך מספרי.
 5. להסיר שורות עם ערכים חסרים בעמודות קרייטיות.
 6. לסנן שורות עם מחירים חריגיים (למשל מעל \$1000 או מתחת ל-\$40).
 7. למלא ערכים חסרים בעמודות משנהות עם ממוצע או חציון.
 8. לשמר את הקובץ cleaned_data.csv.
-

שלב 2: הנדסת מאפיינים (Feature Engineering)

מטרות:

- להפוך עמודות טקסט לייצוג מספרי.
- להכין את הנתונים לקראת מודל ML.

משימות לביצוע:

1. לבדוק אילו עמודות הן קטגוריאליות (כמו סוג החדר).
 2. לבצע קידוד One-Hot לאוותן עמודות.
 3. לוודא שאין בעיות של ערכים חסרים לאחר ההמרה.
 4. לשמר את הקובץ featured_data.csv.
-

שלב 3: חלוקה לאימון ובדיקה

מטרות:

- לחלק את הנתונים ל- Train ו- Test.
- להכין את סט הנתונים למודל.

משימות לביצוע:

1. להפריד בין משתני הקלט (features) למשתנה המטרה (price).
 2. להשתמש בפונקציית train_test_split ולבחור למשל 80% לאימון, 20% לבדיקה.
 3. לשמר את הסטים אם צריך להערכת מאוחרת.
-

שלב 4: בניית מודל AI

מטרות:

- לאמן מודל לחיזוי מחיר.
- לבחור אלגוריתם ולכון פרמטרים בסיסיים.

משימות לביצוע:

1. לבחור אחד מהמודלים הבאים: CatBoost / XGBoost / LightGBM.
2. לאמן את המודל על סט האימון.
3. לשמר את המודל המאומן כ-`apk`. (לא חובה, מומלץ לקרוא על זה)
4. למדוד ביצועים על סט הבדיקה (לא שימוש לעשוות `Cross Validation`)

MAE ○

R^2 ○

5. לרשום הסבר על תוצאות ההערכה: איפה הוא טועה, האם המחיר מנופח מדי?

שלב 5: בניית משק Streamlit

מטרות:

- לאפשר למשתמש להזין פרטי נסס.
- להציג את המחיר הנוכחי.

משימות לביצוע:

1. ליצור טופס קלט:
 - בהתאם למאפיינים שבחרתם במודל.
2. להמיר את הקלט לפורמט שהמודל מבין.
3. להשתמש במודל ולחזות את המחיר.
4. להציג את המחיר בצורה ברורה.
5. להוסיף גרפים או מפה להצגת התפלגות מחירים.
6. לבדוק מה קורה אם מכנים ערכים לא תקינים.

שלב 6: שילוב ML להסברים

מטרות:

- להוסיף רובד של ניתוח בשפה טבעית.

משימות לביצוע:

1. לשלוח בקשה ל- LLM עם המחיר והקלט המקורי.
 2. לבקש הסבר פשוט על המחיר ("למה דואק אכיה?").
 3. להציג את הטקסט זהה בטלגרם או בדשبورד.
 4. בנו דאטאבייס ששמיר את ה- prompt והתשובה ל- prompt המתאים מהמודול.
-

שלב 7: תיעוד והעלאה ל-Github

משימות:

1. לכתוב README שמסביר:
 - מה הפרויקט עושה .
 - דוגמאות לקלט ופלט.