פרוייקט גמר – בנימין וליעד

דגשים:

* הגדרת הבעיה: מה בעייתי ומה אפשר לחזות, קשר בין מאפיינים
* הרכשת נתונים ואיך הרכשנו אותם
* טיפול בנתונים
* EDA ווזיואליזציה
* כל מיני פונקציות ובלאגן של טיפול בנתונים כמו מודל חיזוי וכו
* הערכת ביצועים של המודלים
* הסקת מסקנות ודיווח מסכם/סיכום

מיוחדים:

* פרופיל גיאטהאב + ליצור משתמש וללהעלות את התהליך של ה]רוייקט (תיעוד)

Text

Description automatically generated

רעיון:

התפרצויות הר געש – האם ניתן לחזות אותם, את מיקומם, ואת עוצמתם.

האם יש קשר בין טווחי זמנים מסוימים לעוצמת ההתפרצויות

האם מיקום הר הגעש משפיע על עוצמתו והאם ניתן באמצעות מיקומו לחזות לטווח של מספר שנים את התפרצותו.

האם ניתן לחזות עוצמתו של הר געש על פי סוגו

האם יש קשר לסוגו של הר הגעש לעוצמה של ההתרפצויות

מאפיינים של הרי געש:

VEI – עוצמת ההתפרצות של הר געש

תאריך התחלה – מתי התפרץ

מדינה

LATITUDE

LONGTITUDE

סוג ההר געש – יש 5 סוגים.

גובה ההר - SUMMIT

גובה ההר מעל פני הים - ELEVATION

אזור וולקני – VOLCANIC REGION

דבר ראשון נגדיר מה זה התפרצות?  
התפרצות זה מהרגע שהר הגעש מתפרץ עד שהוא נכבה לחלוטין.

מה הכוונה? למשל הר געש התפרץ וההתפרצות שלו נמשכה מספר ימים או למשל שהר געש מתפרץ בחלקים כל אלו נחשבים להתפרצות אחת – לכן נבדוק כפילויות על פי תאריך תחילת ההתפרצות

אנו מניחים שהר געש התפרץ ביום מסוים והוגדר שוב כהר געש לא פעיל והתפרץ שוב – כל זה באותו היום הוא מקרה שלא קיים.

שלב ראשון: הרכשת הנתונים:

יש את הנתונים שמוצגים ברשימת החיפוש, אנו נחפש את שלושת סיווגי ההתפרצויות:

התפרצויות בטוחות

התפרצויות לא בטוחות

התפרצויות מופרכות

ברשימת החיפוש יש 4 נתונים:

סיווג ההתפרצות (קרה/לא קרה/לא בטוח שקרה)

מדד העוצמה

תאריך ההתפרצות

שם הר הגעש – המון כפילויות.

\*)נבדוק כפילויות ע"פי תאריך ביחס לשם הר הגעש (התייחסות להגדרה של מה היא התפרצות – אין יותר מיום אחד)

**CRAWLING:**

רשימת החיפוש(מה שצוין לעיל)

שאר הנתונים:

מדינה

אזור וולקני

סוג ההר געש – 5 סוגים

גובה ההר summit-

גובה מעל פני היםelevation-

Latitude

Longitude

11 עמודות בסה"כ.

מסמן את כל השורה בטבלהtr -

מסמן את הערכים עצמםtd-

הערות 23/12/21:

דולב שלח לנו שיטות ללחוץ על כפתורים ולינקים עם טיימרים.

להמשיך קרולינג.

יש תיקיות עם מלא לינקים בבוקמרקס.

לאחר הרכשת נתונים:

שני דאטא פריימים:

1. הטבלה של הלינק של תוצאות החיפוש
2. הרשימה עם כל הערכים(גובה,סוג...)

יש למזג בין שניהם בצורה כרונולוגית – כלומר נעשה שרשור של שני הטבלאות לאחת לפי סדר הרשימה (בהנחה והסדר זהה, כלומר השורה הראשונה ממנה הרכשנו את הנתונים המצוינים ב2) נשרשר אליה לפי האיבר הראשון ברשימה)

ניקוי הנתונים:

נרצה להשאיר בדאטאפריים הסופי רק את ההתפרצויות שבוודאות קרו.

נסנן אותם, כל עוד ההתפרצות אינה ודאית – נמחק.