שלום,

שמי מיטל והיום אני אציג לכם את הפרויקט שלי במסגרת הקורס של פייתון, קראתי לפרויקט seenopsis, ואני מאמינה שעוד מעט תבינו למה.

##############

קצת רקע לגבי למה חשבתי על הפרויקט הזה. למעשה הפרויקט הזה בא לתת מענה לצורך בעבודה היומיומית שלי כאפידמיולוגית. רציתי לעשות פרויקט עם תוצר בעל משמעות.

בניגוד למה שכולם חושבים, אפידמיולוגית היא לא מישהי שעוסקת באפידמיות, התפרצות של מחלות, אלא במחקר רפואי שהוא ברמת האוכלוסייה. היום אני עובדת עם מסד הנתונים של שירותי בריאות כללית כדי להפיק תובנות מהדטה, כולל מודלי ניבוי, בחינת אפקטיביות של טיפולים ודיווח של צריכת טיפולים.

\*\*באסכולה שלי, בצורה שבה התחנכתי כאפידמיולוגית, אנחנו תמיד אוהבים להרגיש את הדטה, להבין אותו, להסתכל על ההתפלגויות של הנתונים, על חסרים. כיוון שאני עוסקת בנתונים שהם רפואיים, אנחנו לא יכולים להרשות לעצמנו פשוט לשפוך את כל הדטה למחשב ולראות מה יוצא.

אצלנו תמיד אומרים garbage in garbage out, ולכן יש חשיבות לתת תשומת לב לכל משתנה בנפרד.

אבל להסתכל על כל משתנה זו עבודה סיזיפית, רפטטיבית ו- time consuming.

\*\*לכן חשבתי שכדאי להמציא כלי שנותן אפשרות להסתכל על הנתונים ב- One snap shot, בצורה ויזואלית ויאפשר לי לקבל את כל התובנות שאני צריכה על המשתנים במבט אחד מהיר.

לכן המצאתי את seenopsis!

###############

Seenopsis נותנת אפשרות לעשות סקירה ראשונית של כל המשתנים בדטהסט,

\*\*מרכזת את כל הפיצ'רים החשובים של המשתנים

\*\*והכל כמובן בצורה ויזאולית, קלה לתפיסה.

###############

ואיך בניתי את סינופסיס?

סינופסיס נבנתה בפייתון, וחוץ מהעובדה שזה במסגרת קורס פייתון של shecodes, פיתון היא שפה מאוד נפוצה, שעושים בה המון שימוש ב data-analysis, ולכן זו פלטפורמה מתאימה לצורך שעליו סינופסיס באה לענות.

\*\*ופרט לפייתון, כל מה שצריך כדי להפעיל את סינופסיס זה דטהסט.

\*\*אז כמובן, אין טוב ממראה עיניים, בוא נלך לסינופסיס ונראה איך היא פועלת.

###############

בואו נסתכל על הארכיטקטורה של סינופסיס וקצת נבין מה עבר לי בראש.

אז במקור, ייעדתי את סינופסיס לדטה פריים שך pandas שזו חבילה נפוצה בפייתון ל- data analysis. אבל אז קלטתי שאצלנו במכון למשל השליפות נעשות ב- SQL, והאנליזה לרוב נעשית בכלל ב- R,

\*\*ולכן חשבתי שצריך לתת אפשרות לקרוא את הקובץ גם בפורמט שהוא יותר גנרי, מחוץ לפייתון, והוספתי את האופציה של קריאת CSV.

כמובן שזה לא עבר ללא תקלות, מסתבר שתוכנות שונות מייצאות קבצי CSV תוך שימוש ב- encoding שונה. כרגע seenopsis יודעת לקרוא קבצי CSV שמופקים בencoding הנפוצים. הרצתי את סינופסיס על הרבה קבצי CSV שונים, ולא נתקלתי בתקלה של encoding, אני מניחה שבעתיד יהיה הקובץ שיש לו encoding מיוחד, ולכן בניתי את הקוד בצורה כזו שבקלות אפשר להוסיף את סוג ה- encoding ולהצליח לקרוא אותו.

\*\*

ברגע שיש דטהסט, צריך לייבא את סינופסיס ולהריץ אחת משתי הפונקציות כתלות בסוג הדטהסט:

או Seenopsis.process\_csv – ואז נפתח חלון ב- browser שבעזרתו אפשר לבחור את המיקום של הקובץ.

או seenopsis.process\_pandas\_df – שמאפשר לעבוד ישירות מפייתון. רק צריך לספק לו את השם של ה- dataset.

\*\*

ברגע שמריצים את סינופסיס, כפי שראיתם חלק מה- output של סינופסיס זה הצגה גרפית של המשתנים. וכדי להציג את זה בממשק ה- output שבחרתי, היה צורך לייצר את כל הגרפים הללו באופן פיזי, ולכן בכל הרצה של סינופסיס, תיווצר תיקיה בסביבת העבודה של סינופסיס, בשם “xxxxx”, ובתוכה יהיו כל הגרפים של הקובץ. הכל נוצר באופן אוטומטי.

\*\*

בסוף, ה- output מיוצר כקובץ html, זה היה הפיתרון הכי גנרי שיכולתי לחשוב עליו, ושאפשר גם לשחק איתו מבחינת ויזואליזציה של הנתונים, וגם על הדרך למדתי קצת על בנייה של קבצי html, ונגעתי קצת ב- front end, אז זה win win situation. פה קיבלתי קצת עזרה מאחי, הצגתי את הפרויקט למשפחה, ובהתחלה ה- output היה טבלה יחסית פשוטה, ואחי שמסתבר יודע קצת front end, עזר לי בליצור טבלה שאפשר לגלול, כשהכותרת נשארת קבועה. זה היה מאוד נחמד לעבוד עם אחי על פרויקט מקצועי.

########

בויז'ן שלי, סינופסיס צריכה להיות library שמייבאים כמו כל שאר ה- libraries בפיתון, ובויז'ן שלי, כל מי שמתעסק עם דטה, בלי חשיבות לדומיין, הדבר הראשון שהוא יעשה זה להריץ את סינופסיס כדי לדעת מה הם המשתנים שהוא עובד איתם.

ואני באמת מאמינה שסינופסיס נותנת מענה מהיר וחשוב באנליזה של כל מסד נתונים.

#########

אז כמובן, שיש עוד פיצ'רים עתידיים שאני רוצה לפתח לסינופסיס.

\*\*חשבתי לעשות מזה איזשהו web application, שלא יהיה סגור רק למשתמשי פייתון. אני אעשה את זה אחרי שאני אעשה קורס web ב- shecodes.

\*\*עשיתי איזשהו ניסיון לייצר PDF אוטומטי כדי שאפשר יהיה בקלות להעביר את הseenopsis של הדטה לחוקרים חיצוניים, לא הספקתי לעשות את זה במסגרת הזמן. מצאתי פיתרון לזה, אבל הוא היה קצת מורכב, אז החלטתי לא לשחרר את הפיצ'ר הזה בגרסה הזו.

\*\*כרגע, סינופסיס עובדת באופן מקומי, כך שיש גבול לגודל הקבצים שאפשר להעמיס, בתלות במחשב המקומי. יהיה נחמד להעביר את הכח האנליטי של סינופסיס לאיזה שרת שיוכל להתמודד עם קבצים גדולים.

\*\*הקובץ HTML, נראה מאוד יפה בכרום, אבל הגלילה לא עובדת ב- webbrowsers אחרים, אז זה עוד משהו ששווה לשפר.

\*\*כרגע בכל הרצה, נוצרת תיקיה של גרפים, וזה תופס מקום. צריכה לחשוב על פיתרון יצירתי אחר שבו לא צריך ליצור פיזית את הגרפים הללו, אלא אולי רק שיווצרו באופן טמפורלי וימחקו.

################

אז כל התהליך הזה היה תהליך למידה.

חוץ מזה שלמדתי המון פייתון, כי אין דרך יותר טובה ללמוד משהו מאשר לעסוק בו, למדתי גם

\*\* שמאוד חשוב לדבר עם אנשים, עם קולגות, משפחה, סתם חברים, לכולם יש רעיונות טובים, ואלו הם בעצם משתמשי הקצה שאליהם צריך להגיע. ולכן חשוב מאוד לדבר עם אנשים ולשתף.

\*\*חוץ מזה, ניסיתי את סינופסיס כמעט על כל קבצי ה- CSV שיש ב- kaggle, מכל קובץ מצאתי באג ושיפרתי את הקוד שיהיה יותר generalized, כל קובץ נתן לי מקרה אחר של real world.

\*\*ובסוף, הבנתי שצריך תשוקה למה שעושים. יש לי שלושה ילדים, עבודה במשרה מלאה ובעוד שבוע אני עוברת דירה, ועדיין, הצלחתי להרים את הפרויקט. בעיקר כי מאוד נהנתי מהתהליך, ולכן הצלחתי לגייס אומץ ואמונה שאני יכולה לעשות את זה ואופטימיות שזה יצליח.

אז היה ממש ממש כיף!

תודה על ההזדמנות!