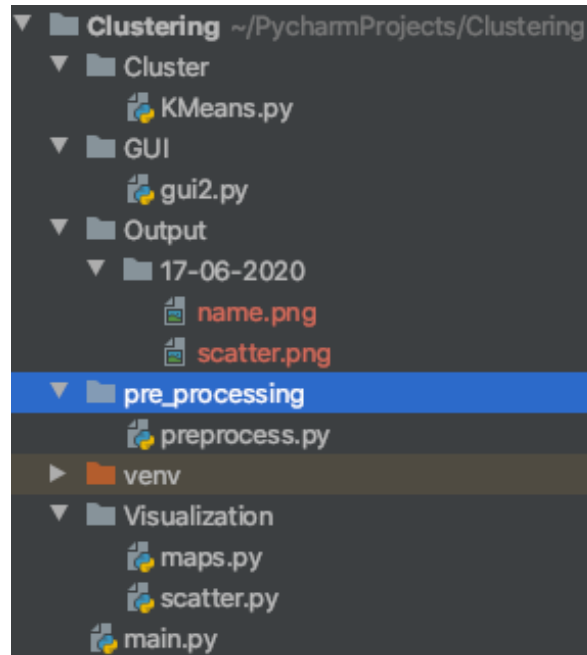


# דו"ח עבודה 4

יפתח סברנסקי 312141369

שירה ורטהיים 313546442

מצ"ב המחלקות והחבילות בתוכנית:



function description	function	class description	class name	package description	package
k_means_model - בונה מודל clustering בעזרת אלגוריתם k-Means ומחזיר ווקטור המכיל את החלוקה לקלאסטרים עבור כל רשומה.	k_means_model(data_frame, num_clusters, num_init:)	- Kmeans המחלקה אחראית על ביצוע Clustering בעזרת אלגוריתם k-means.	KMeans	Cluster - החבילה מכילה מחלקות שאחראיות על ביצוע Clustering באמצעות אלגוריתמים שונים (כרגע מכילה מחלקה יחידה).	Cluster
k_means_cluster - מקבל dataframe לחלק את הרשומות בו לקלאסטר, את מספר הקלאסטרים הרצוי ואת מספר ההרצות הרצויות של המודל. מחזיר את dataframe שהתקבל בצירוף העמודה 'cluster' המכילה את החלוקה לקלאסטרים ע"פ המודל.	k_means_cluster(data_frame, num_clusters, num_init:)				
אתחול המחלקה	__init__(self, *args, **kwargs:)	מחלקה אשר אחראית לתפעול השוטף של הGUI בתוכנית.	gui2	החבילה מכילה מחלקות שאחראיות על ביצוע ממשק GUI באמצעות אלגוריתמים שונים	GUI
אתחול החלון הראשי GUI	InitUI(self:)				
אתחול הכפתורים והפונקציונאליות בGUI	InitMainPanel(self:)				
אתחול מתן הנתיב לקובץ excel, וחלונות למתן מספרים של clusters.	loadWidgets(self, panel:):				
פונקציית עזר השומרת את הנתיב לקובץ excel	on_open_file(self, event:):				
פונקציית עזר השומרת את מספר ה clusters	catchClusterNumber(self, event:):				
פונקציית עזר השומרת את מספר ה runs	catchRunNumber(self, event:):				
קריאה לפונקציה pre_process() מהmain של התוכנית. זריקת שגיאה בחלון מותאם במקרה והנתיב לא תקין, או הקפצת התראה הצלחה בחלון מתאים.	preProcessFunc(self, e):				
קריאה לפונקציות getDataFrameLength() ולcluster() מהmain של התוכנית. זריקת שגיאה בחלון מותאם במקרה והמספרים לא תקינים. בסוף הפונקציה מדפיס לחלון הGUI את 2 הפלטים הרלוונטיים. כמו כן דואג להקפצת התראה הצלחה בחלון מתאים.	clusterFunc(self, e):				
תחילת ביצוע הפקודות ליצירת הGUI	main:()				
"Reading a csv dataset into a Pandas Dataframe ":param path: path to a csv returns: Pandas Dataframe :" with the dataset from the file	read_csv(path:)	המחלקה מכילה את השלבים הנדרשים לביצוע בשלב העיבוד המקדים של הנתונים עבור העבודה הנוכחית.	preprocessor	החבילה מכילה מחלקות שונות שימשו לטובת ביצוע עיבוד מקדים של נתונים.	pre_processing
"Reading a dataset from an excel into a Pandas Dataframe ":param path: path to a excel file containing the desired dataset returns: Pandas Dataframe :" with the dataset from the file	read_excel(path:)				

<p>Receives a dataframe and fills all the missing values in the numeric columns with the mean of each column.</p> <p>:param data_frame: a dataframe with numeric columns containing missing values</p> <p>returns: a dataframe with no missing values in the numeric columns</p>	fill_missing_values_numeric(data_frame:(				
<p>Receives a dataframe and standardize all the values in the numeric columns.</p> <p>:param data_frame: a dataframe with numeric columns containing no missing values</p> <p>returns: a standardize : " dataframe</p>	z_score_standardization(data_frame:(				
<p>Receives a dataframe and groups all the years into a single entry. Calculates the mean of each attribute across the years and places it as the new attribute value,</p> <p>:param data_frame: a dataframe with 'country' and 'year' columns and numeric columns containing no missing values</p> <p>returns: a dataframe grouped : " by country</p>	group_by_country(data_frame:(				
<p>מבצעת את השלבים הנדרשים לביצוע שלבי העיבוד המקדים של הנתונים עבור העבודה הנוכחית. פרמטר normalize קובע האם יתבצע נירמול לטווחי הנתונים.</p>	preprocess(path, normalize:(				
				התיקיה מכילה את הפלט של הפעולה. מכילה את הגרפים שיוצאו.	Output
<p>מקבלת dataframe (המכילה עמודות של 'country' ו-'cluster') ונתיב לפלט ושומרת לנתיב הפלט מפה הצבוע לפי הקלאסטרים</p>	plot_choropleth_map(data_frame, output_path:(	המחלקה אחראית על הצגת נתונים על מפה גיאוגרפית	maps	החבילה מכילה מחלקות שמטרתן היא להציג נתונים בצורה חזותית וברורה.	Visualization
<p>שומרת תרשים מפה כתמונה בנתיב הפלט שהתקבל כפרמטר</p>	save_choropleth_map(choromap, output_path:(				
<p>מקבלת את הנתונים הנדרשים ושומרת גרף פיזור בנתיב הפלט שהתקבל</p>	plot_scatter(data_frame, x_name, y_name, x_label, y_label, title, color_by_name, output_path:(	המחלקה אחראית על הצגת נתונים על גרף כפיזור נקודות.	scatter		
<p>מבצעת את פעולת הPre-processing.</p>	pre_process(pathFromGui):	המחלקה אחראית לריצת התוכנה וכוללת בתוכה פונקציות המרכזות בתוכן את הלוגיקה של ההרצה.	main		
<p>מחזירה את אורך dataframe הטעון בזיכרון</p>	getDataFrameLength():				
<p>מבצעת את החלוקה לקלאסטרים באמצעות אלגוריתם k-means ושומרת תצוגה גרפית של תוצאות החלוקה</p>	cluster(clusterInput, runInput):				

מתחילה את ריצת התוכנה (מפעילה את ממשק המשתמש)	start():				
מטרת פונקציה זו היא לבצע ניתוח של הקלאסטרים ולהבין מה המאפיינים של כל קלאסטר (ביצוע ממוצע על ערכי כל פיצ'ר לכל הרשומות מאותה קבוצה) - לא משומשת (אבל מעניינת)	clusterGrouping():				