



תרגיל בית 2

סט הנתונים שנעבוד איתו בתרגיל נתון במודל תחת השם "dpc-covid19-ita-region.csv".

סט הנתונים, בעל כ-7800 רשומות, מתאר את נתוני התחלואה במחוזות איטליה בין ה-24/02/2020 ל-28/02/2021 (371 ימים). סט הנתונים מתועד ומתוחזק ע"י רשות החירום הלאומית של ממשלת איטליה תחת הלינק:

<https://github.com/pcm-dpc/COVID-19>

נעבוד בתרגיל עם חלק מהתכונות (עליכם לעבוד רק עם הקובץ הנתון במודל). התכונות שנעבוד עימן הן:

Field name	Description	Format	Example
date	Date of notification	YYYY-MM-DDTHH: MM: SS (ISO 8601) Italian time	2020-03-05T12: 15: 45
region_code	Code of the Region (ISTAT 2019)	Number	13
denominazione_region	Name of the Region	Text	Abruzzo
hospitalized_with_symptoms	Hospitalized patients with symptoms	Number	3
Intensive_care	Intensive Care	Number	3
total_hospitalized	Total hospitalized patients	Number	3
home_insulation	Home confinement	Number	3
new_positives	News amount of current positive cases (total_cases current day - total_cases previous day)	Number	3
resigned_healed	Recovered	Number	3



תיאור המשימות

נגדיר:

- ממוצע:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i$$

- חציון:

חציון הוא הערך שמחצית מהתצפיות קטנות או שוות לו, ומחצית מהתצפיות גדולות או שוות ממנו. יהי n מספר התצפיות. אזי, אם n זוגי:

$$Med = \frac{X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)}}{2}$$

אחרת:

$$Med = X_{(\lfloor \frac{n}{2} \rfloor)}$$

כאשר $X_{(i)}$ הינו האיבר ה- i ברשימת הערכים הממוינת בסדר לא יורד.

שאלה 1:

בשאלה זו נתעניין בקבוצת המחוזות D . המחוזות D הם כל המחוזות ששמן (denominazione_region) מתחיל באותיות 'S' או 'L'.

עליך לחשב עבור המחוזות בקבוצה D ממוצע וחציון עבור התכונות:

- hospitalized_with_symptoms
- Intensive_care
- total_hospitalized
- home_insulation

פלט משימה זו צריך להיות מהצורה:

Question 1:

hospitalized_with_symptoms: value of mean, value of median

intensive_care: value of mean, value of median

total_hospitalized: value of mean, value of median

home_insulation: value of mean, value of median



שאלה 2:

בשאלה זו נעזור לקבינט הקורונה האיטלקי להחליט על מדיניות הסגרים האיטלקית. עליכן לפתור ויכוח בין אדון גמזוסו, ראש שירותי בריאות הציבור האיטלקי, לבין אדון אנתוניו, ראש קבינט הקורונה האיטלקי.

נגדיר רשומה כרשומת "יום טוב למחוז" באם מספר המחלימים בה היה גבוה ממספר המאובחנים החדשים. כלומר, רשומה x_i תיחשב "יום טוב למחוז" באם $\text{resigned_healed_i} - \text{new_positives_i} > 0$, ואחרת תיחשב "יום רע למחוז".

נגדיר מחוז כ"מחוז ירוק", אם מספר הימים הטובים שתועדו עבורו בנתונים גדול מ340.

מר גמזוסו טוען שעל המדינה להטיל סגר על מחוזות שאינם ירוקים בלבד. מאידך, מר אנתוניו טוען שיש להטיל סגר על כל המדינה.

למזלנו, לאחר דיונים ממושכים, מר גמזוסו ואנתוניו הגיעו להסכמה שאם מספר המחוזות שאינם ירוקים המתועדים בנתונים גדול מ10, יוטל סגר על כל המדינה, ואחרת יוטל סגר רק על המחוזות שאינם ירוקים. הסכמה בין השניים תקבע את מדיניותה של ממשלת איטליה.

עליכן לקבוע - האם יוטל סגר על כל איטליה או לא?

פלט משימה זו צריך להיות מהצורה:

Question 2:

Number of districts: Number of districts

Number of not green districts: Number of not green districts

Will a lockdown be forced on whole of Italy?: Yes/No

דוגמה לפלט שמבוסס על מדגם קטן ושרירותי של תצפיות מתוך סט הנתונים:

```
Question 1:
hospitalized_with_symptoms: 1323.8429319371728, 472
intensive_care: 139.89005235602093, 47
total_hospitalized: 1463.7329842931938, 515
home_insulation: 15320.4554973822, 6396

Question 2:
Number of districts: 21
Number of not green districts: 21
Will a lockdown be forced on whole of Italy?: Yes
```

מבנה הפלט צריך להראות בדיוק על-פי הדוגמה המצורפת, מאחר ומתבצעת השוואת קבצים אוטומטית.

שימו לב: המדגם הנ"ל נתון לכם תחת הקובץ `dpc-covid19-ita-regioni_sample.csv`.



דרישות למימוש:

המימוש **חייב** להכיל לפחות את המודולים, המחלקות והמתודות הבאות:
ניתן להוסיף data members למחלקות המתוארות.

1. **main.py** – ממשק ראשי לריצת התוכנית, כפי שהיא מתוארת בפרק "תיאור המשימות". בין היתר, בקובץ זה יופיעו השורות:

```
def main(argv):  
    pass  
  
if __name__ == '__main__':  
    main(sys.argv)
```

כאשר במקום "pass" יופיע קטע הקוד. עליכן לייבא את הספריה sys בראש הקובץ, ע"י import sys. ע"י המבנה הנ"ל תוכלו להריץ את התוכנית שלכם באמצעות:

python **your_path**/main.py arguments

בתרגיל בית זה:

python /home/student/**your_path**/main.py /home/student/**your_path**/dpc-covid19-ita-regioni.csv

שימו לב:

argv[0] = /home/student/**your_path**/main.py,
argv[1] = /home/student/**your_path**/dpc-covid19-ita-regioni.csv

כאשר **your_path** הוא הנתיב בו שמור המודול main.py, שאר המודולים, וקובץ הנתונים dpc-covid19-ita-regioni.csv.

(15%)

2. **data.py** – ממשק לאיסוף הנתונים.

מחלקה Data:

בנאי המחלקה __init__ מקבל כקלט את path (הנתיב המלא של קובץ הנתונים).

Data members:

data - מילון שהkeys שלו הם תכונות מסט הנתונים, והvalues הם רשימות שמכילים את ערכי התכונות.

עליכן לבנות את data באותה צורה בדיוק כמו בתרגיל בית 1. לנוחיותכם, קטע הקוד המתאים:

```
df = pandas.read_csv(path)
```



```
self.data = df.to_dict(orient="list")
```

לצורך כך יש להשתמש בספרייה pandas, (על ידי Import pandas).

מתודה:

get_all_districts(self)

המתודה מחזירה את רשימת שמות כל המחוזות השונים בdata (מחוזות denominazione_region).
(5%)

מתודה:

set_districts_data(self, districts)

פרמטרים:

- districts הינה רשימה של מחוזות (denominazione_region).
תפקיד המתודה הוא לעדכן את השדה data כך שיכיל את המחוזות ששם נמצא בתוך הרשימה districts.
כלומר, לאחר הפעלת הפונ' המילון data יכיל את כל הרשומות המתאימות למחוזות ב-districts, ורק הן בלבד.
(10%)

3. districts.py

מחלקה Districts:

בנאי המחלקה __init__ מקבל כקלט אובייקט מסוג Data.

Data members:

dataset – זהו האובייקט מסוג Data שמתקבל כקלט ע"י בנאי המחלקה.

מתודות:

filter_districts(self, letters)

המתודה מקבלת כקלט את letters – set של אותיות. תפקיד המתודה הוא לעדכן את השדה data של האובייקט dataset כך שיכיל רק את המחוזות ששם מתחיל באותיות הנמצאות בletters. שימו לב שעליכם להשתמש במימוש המתודה filter_districts במתודות שמישתם במחלקה Data, ע"י קריאה להן מתוך האובייקט dataset.

print_details(self, features, statistic_functions)

המתודה מקבלת כקלט את features שזהו list של תכונות, ואת statistic_fucntions שזהו list של מתודות סטטיסטיות מהמודול statistics. המתודה מדפיסה מדדים סטטיסטיים על השדה data של האובייקט dataset, אך ורק לפי התכונות בfeatures, תוך שימוש במתודות שב - statistic_functions.



determine_day_type(self)

תפקיד המתודה הוא להוסיף key לשדה data של האובייקט Dataset בשם "day_type", כאשר הvalue הוא list של אינדיקטורים המסמלים לכל רשומה האם היא "יום רע" או "יום טוב" כפי שהוגדר בשאלה 2. במידה והרשומה הינה "יום רע", הערך המתאים לה ברשימה יהיה 0, ואחרת הוא יהיה 1.

get_districts_class(self)

פלט המתודה יהיה מילון שהkeys שלו יהיו שמות כל המחוזות השונים בשדה data של האובייקט dataset. values יהיו המחוזות 'green' או 'not green' בהתאם להגדרה של מחוז ירוק בשאלה 2.

(60%)

4. **statistics.py** – מודול לחישוב מדדים סטטיסטיים.

מתודות שעליכם לממש:

mean(values)

median(values)

פרמטרים:

- values – רשימה של ערכים מספריים.

פלט המתודות הנ"ל הוא כפי שהוגדר בפרק "תיאור המשימות" (כאשר פלט המתודות mean and median הינם הממוצע והחציון בהתאמה). ניתן להשתמש בפונ' ceil של הספריה math לטובת חישוב ערך עליון במתודה median. **חל איסור להשתמש בספריה math להפעלת מתודה אחרת, חל איסור להשתמש בספריות חיצוניות למימוש שאר המתודות במודול הנוכחי.**

(5% למתודה)



דגשים נוספים:

1. עליכם לכתוב את הקוד בהתאם לדגשים והסטנדרטים לפי pep8. לשימושכם המסמך Code Quality Requirements באתר ה-moodle של הקורס. קוד אשר לא יעמוד בסטנדרטים הנדרשים, יקבל ניקוד מופחת.
2. ניתן להוסיף מתודות נוספות, במידה ותמצאו לנכון. יש להימנע מכפילויות קוד.
3. ניתן להשתמש במתודות שהן in-built בשפה. קרי, מתודות אשר לא דורשות ייבוא של ספריות.
4. יש לתת שמות בעלי משמעות לכל משתנה.
5. חובה לתעד את הקוד באנגלית. בפרט עליכם לכתוב עבור כל מתודה docstring.

הוראות הגשה:

- התרגיל להגשה בזוגות בלבד.
 - לפני ההגשה, חובה לוודא שהתוכנית עובדת במעבדת ההוראה ולא בסביבה אחרת.
 - ההגשה חייבת להכיל קובץ אחד (קובץ zip) :
 - שם הקובץ חייב להיות hw2_xxxxxxxxx_yyyyyyyy.zip כאשר xxxxxxxx ו-yyyyyyy הם מספרי תעודות הזהות של המגישים, כולל ספרת ביקורת.
 - הקובץ מכיל את כל קבצי הקוד. אין להכיל תיקייה ובתוכה קבצי הקוד, אלא את קבצי הקוד עצמם.
 - **הערה:** עליכם לוודא שהתוכנית מתחילה לפעול מקובץ "main.py" בלבד.
 - תשובות לחלקים יבשים יש להקליד במעבד תמלילים. אין להגיש תשובות בכתב יד.
 - ההגשה היא אלקטרונית בלבד, דרך אתר ה-moodle של הקורס. תרגילים שיוגשו בכל דרך אחרת לא ייבדקו.
 - אין להגיש את אותו הקובץ פעמיים. התרגיל יוגש ע"י אחד מבני הזוג.
 - שימו לב שההגשה תיחסם בדיוק בשעה 23:55 ביום ההגשה. מומלץ להגיש לפחות שעה לפני המועד האחרון.
 - ניתן להגיש כמה פעמים. רק ההגשה האחרונה תישמר.
 - תרגיל בית שלא יוגש לפי הוראות ההגשה – לא ייבדק (כלומר יקבל ציון 0).
 - לצורך תרגיל הבית ייפתח פורום. ניהול שאלות ומתן תשובות בנושא התרגיל יתבצע דרך הפורום בלבד.
- בהצלחה!