



אנליזה של ביג דאטה תרגיל 1



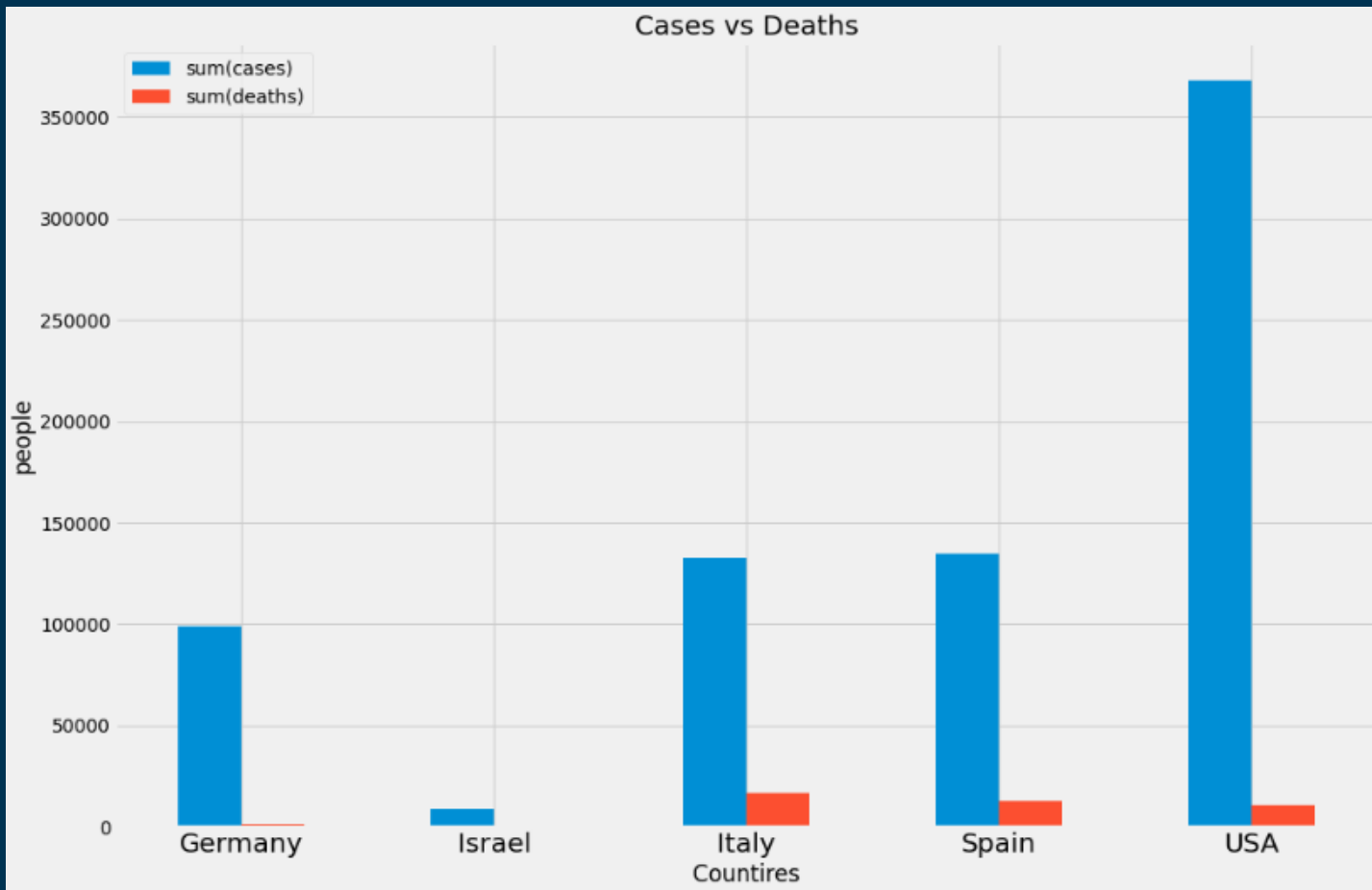
בחירת הדאטה

בתרגיל זה התבקשנו להתרכז בדאטה הקשור למחלת הקורונה והתפרצותה. נציג ניתוח של נתונים אלה ע"י גרפים ומודל רגרסיה לינראית.

לשם כך בחרנו להתמקד בנתונים הבאים :

- נתוני ה-ECDC העוסקים בהתפתחות המגפה בעולם.
- דרכי התמודדות של מדינות למגפה מאת אוניברסיטת John Hopkins.
- נתוני ששוק ההון הפתוחים לציבור, בפרט במדדי הנפט וחברות התעופה.
- במדדי רווחה ואושר של מדינות אירופה המתפרסם מדי שנה.
- מידע על תוחלת החיים של כלל מדינות העולם.

Number of confirmed cases and deaths due to COVID-19

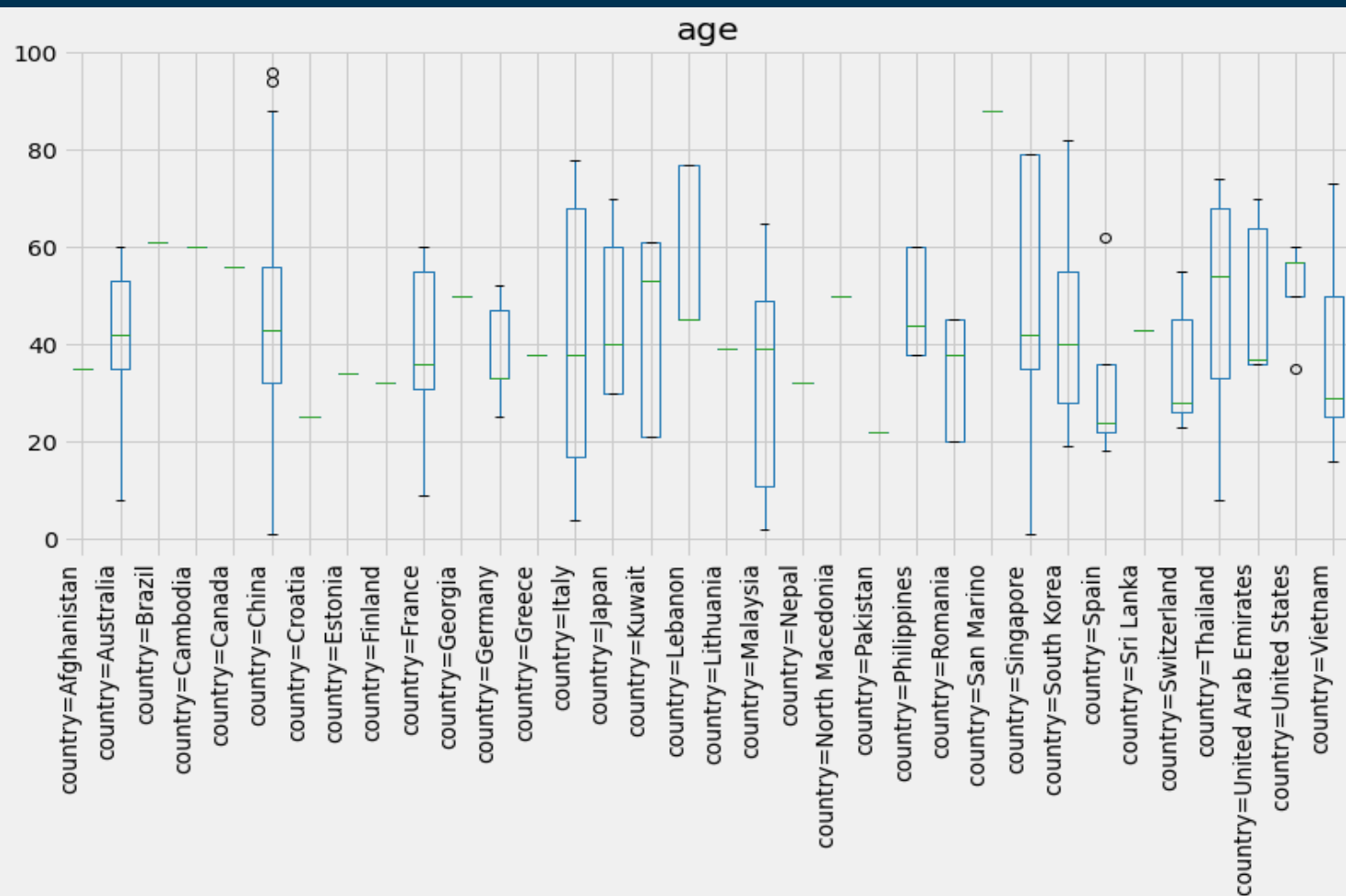


כאן אנו רואים בצורה ברורה את כמות החולים והמתים של מספר מדינות וביניהן ישראל.

נכון להיום אפשר לראות שבכמות הנדבקים בלבד, ארה"ב כרגע נמצאת הרבה מעל כל המדינות. יכול להיות שרק בגלל כמות האנשים, אבל גם אולי יש עוד מדדים המשפיעים על הנתונים הללו.

ניתן לראות גם שכמות המתים של איטליה וספרד גבוהות במיוחד, ואף עוברות את כמות המתים של ארה"ב, למרות שגודל המדינות קטן בהרבה.

How age is distributed for each country



גרף המתאר התפלגות גילאים בשנים
לפי מדינות.

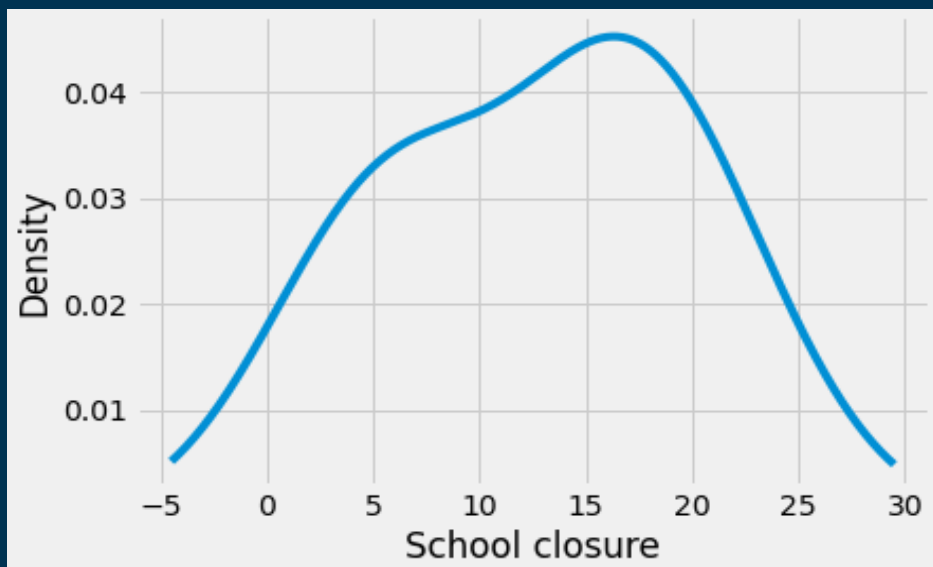
כל עמודה מתארת התפלגות נורמלית
עבור הגילאים של החולים באותה
מדינה.

נתונים אלו יעזרו לנו לבדוק קורלציות
מול נתוני המחלה.

Are we taking precautions?

אם רוב המדינות לא סגרו את בתי הספר(כלומר ערך 0 ברשומה) אז הפיק של הגרף יהיה בנקודה 0.

אם קצת מהמדינות סגרו את בתי הספר(כלומר ערך גדול מ0 ברשומה) אז יהיה עוד פיק בגרף בנקודה ימנית לנקודה 0.



Graph 2
Density vs school closure (over the last day)

מהי פונקצית צפיפות בהקשר של סגירת בתי ספר?

אפשר לחשוב על הפונקציה שהגרף מתאר כגל.

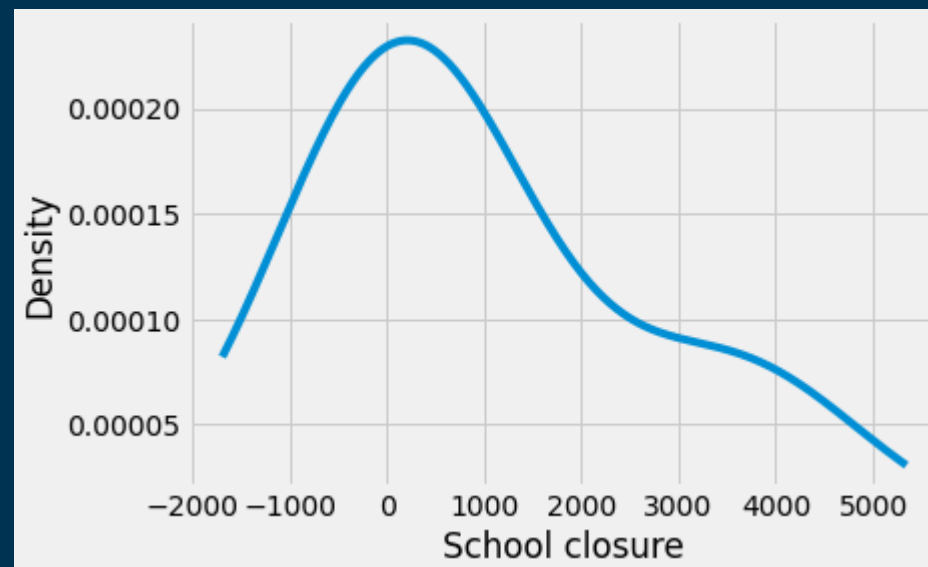
כתגובה כלומר הוא בא לענות על השאלה האם עוד מדינות הולכות לסגור בתי ספר למגפת הקורונה

צפיפות של סגירת בתי ספר ע"י מדינות:

בגרף הראשון אנו מסתכלים על נקודת האמצע מבחינת תאריכים

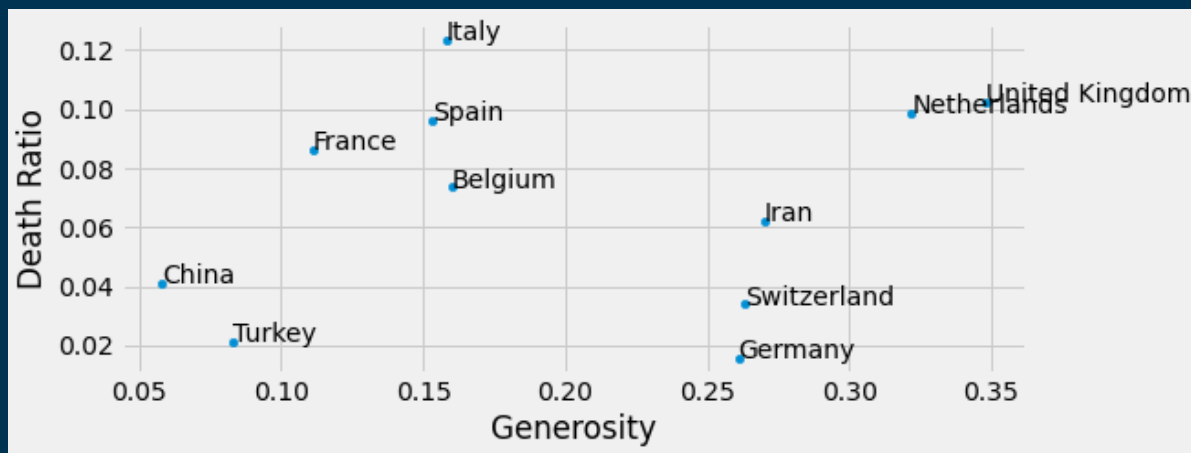
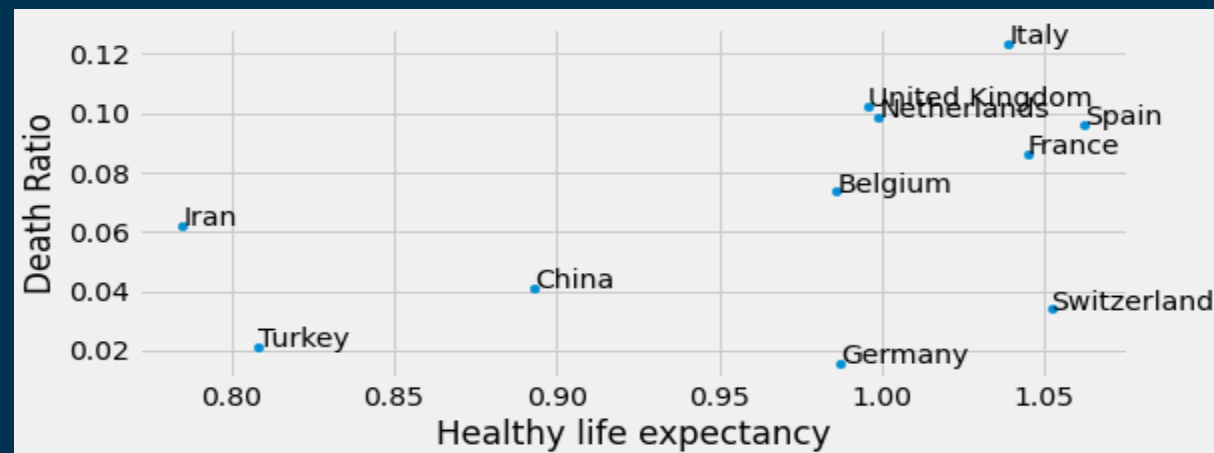
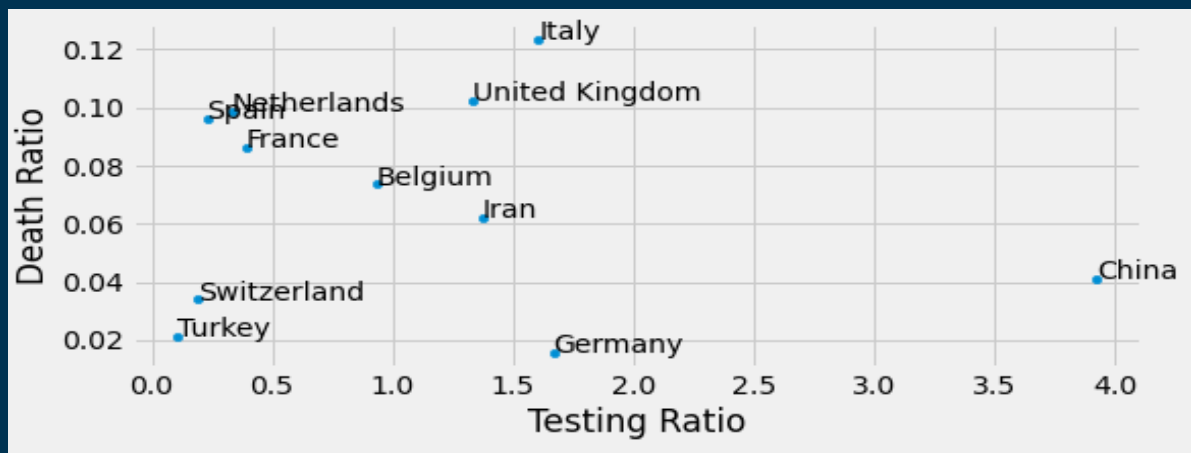
בגרף השני אנו מסתכלים על היום הכי עדכני עבור כל המדינות

ניתן לראות כי לאורך זמן יותר מדינות סגרו את בתי הספר שלהן ככל שהמחלה מתפשטת



Graph 1
Density vs school closure density (over the middle day of the outbreak)

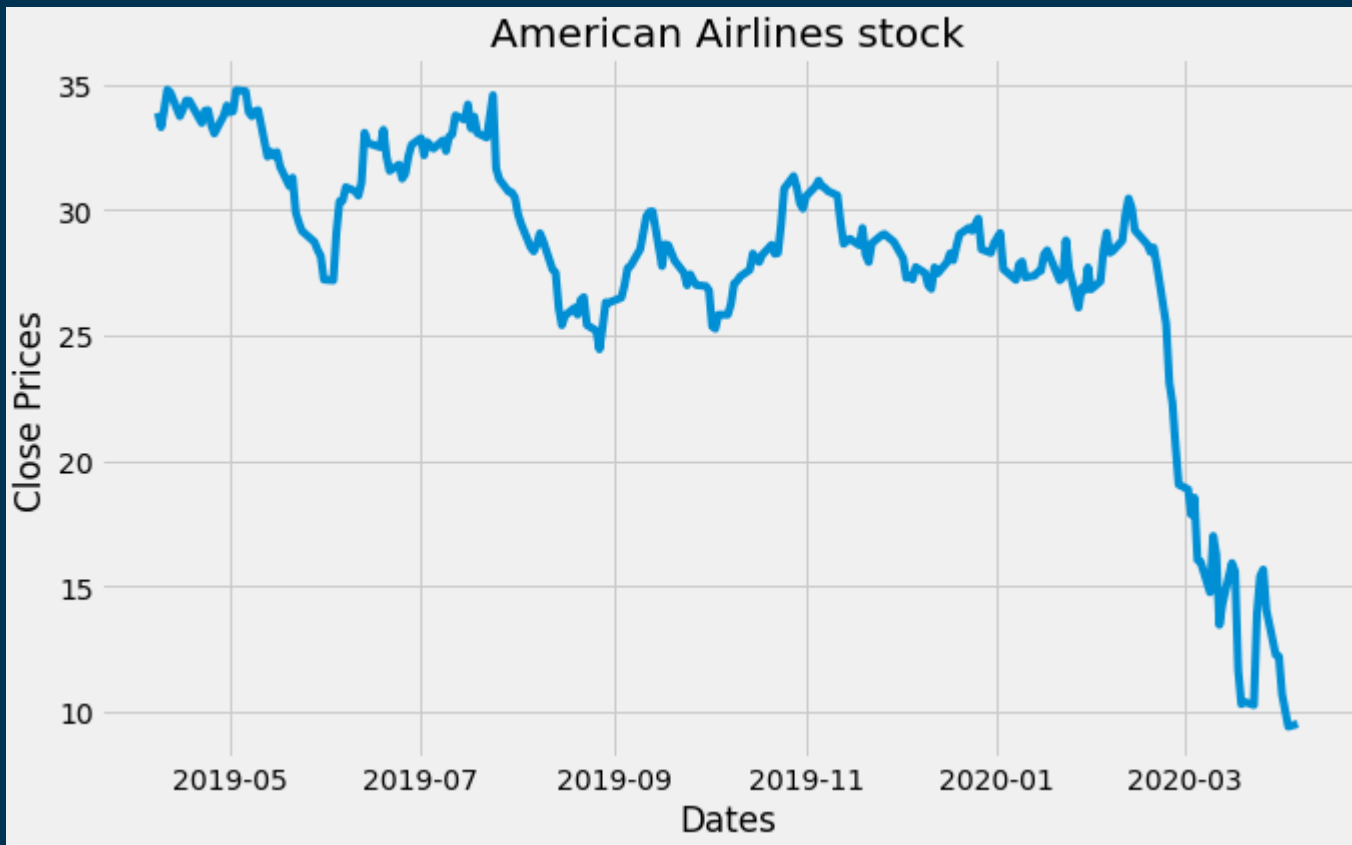
How does quality of life affect death rates



הגדרות

נגדיר את יחס המוות להיות כמות המתים חלקי כמות החולים
נגדיר את יחס הבדיקות להיות כמות הבדיקות חלקי כמות החולים
מדדי האושר של מדינה נקבעים ע"י דירוג של אזרחי המדינה
המדדים כוללים בין השאר צפייה לחיים ארוכים ונדיבות
ניצור גרפים המשווים בין יחס המוות לפרמטרים הבאים
יחס הבדיקות
נדיבות
צפי לחיים ארוכים

COVID-19 effects on the economy



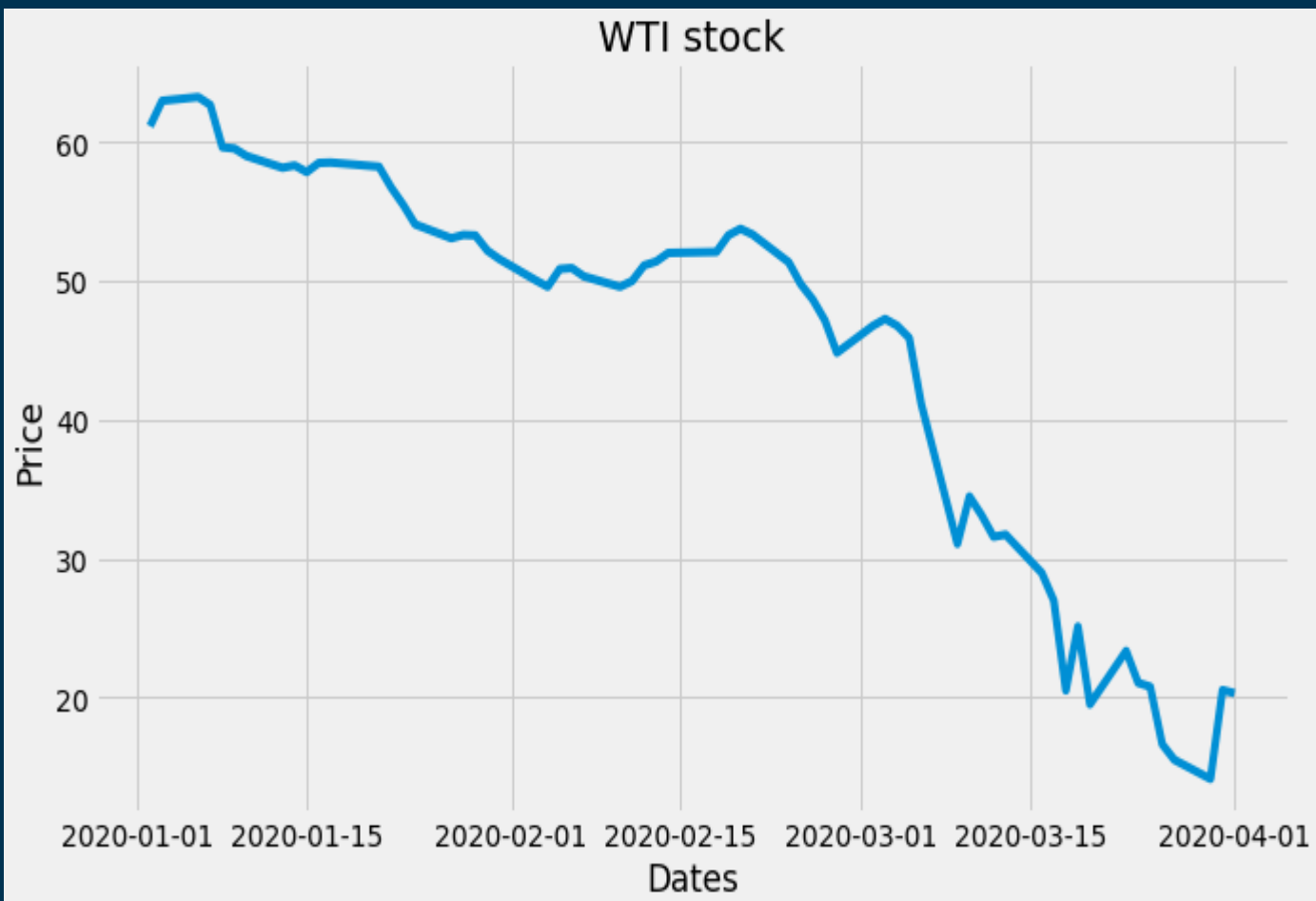
כאן ניתן לראות את שערי הסגירה של מניית AAL בדולרים על פני השנה האחרונה

המשבר שמטביע את תעשיית התעופה העולמית מחרף.

אמריקן איירליינס הודיעו על ירידה צפויה בהכנסות, קיצוצים חדים בטיסותיהן ומהלכי חירום נוספים. זאת, במטרה לצמצם את נזקי נגיף הקורונה שמדלד באופן דרסטי את תנועתה נוסעים העולמית. נכון ל־2019, אמריקן איירליינס היא חברת התעופה הגדולה בעולם עם צי של 956 מטוסים.

ניתן לראות ירידה חדה במיוחד בשערי המנייה בתקופת סוף פברואר. בסביבות אותה תקופה, ארה"ב התחילה לראות עלייה בכמות החולים.

COVID-19 effects on the economy

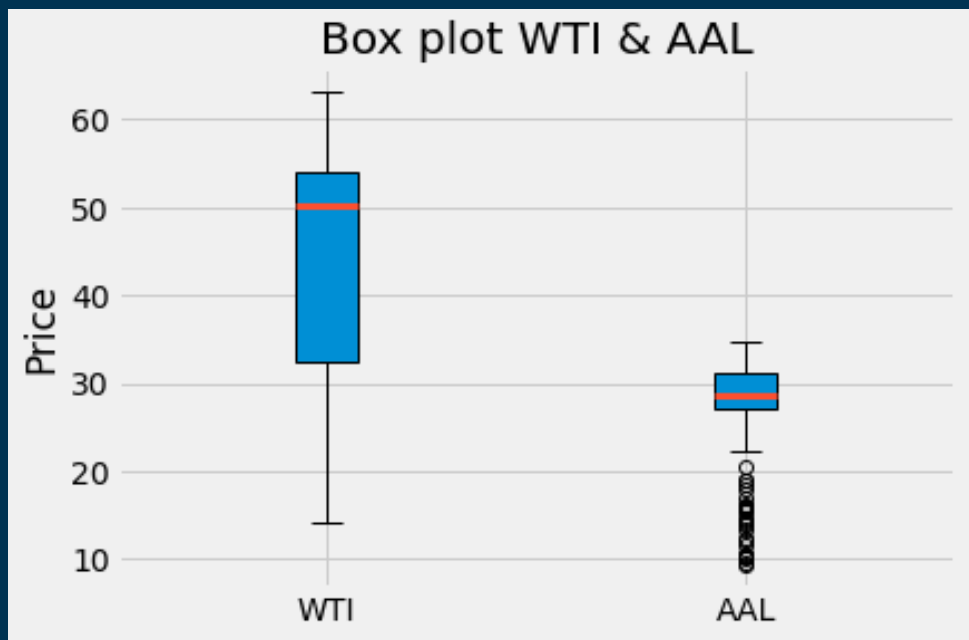


כאן ניתן לראות את שערי הסגירה של מניית WTI בדולרים על פני שנת 2020

בדומה למה שראינו קודם עם מניית אמירקן איירליינס, יש ירידה משמעותית במחיר מאז תחילת המשבר.

לאורך שנת 2020 המנייה ירדה מסביבות 60 דולר ל-20 דולר עד לאפריל. ירידה של כ-66% במשך רבעון אחד.

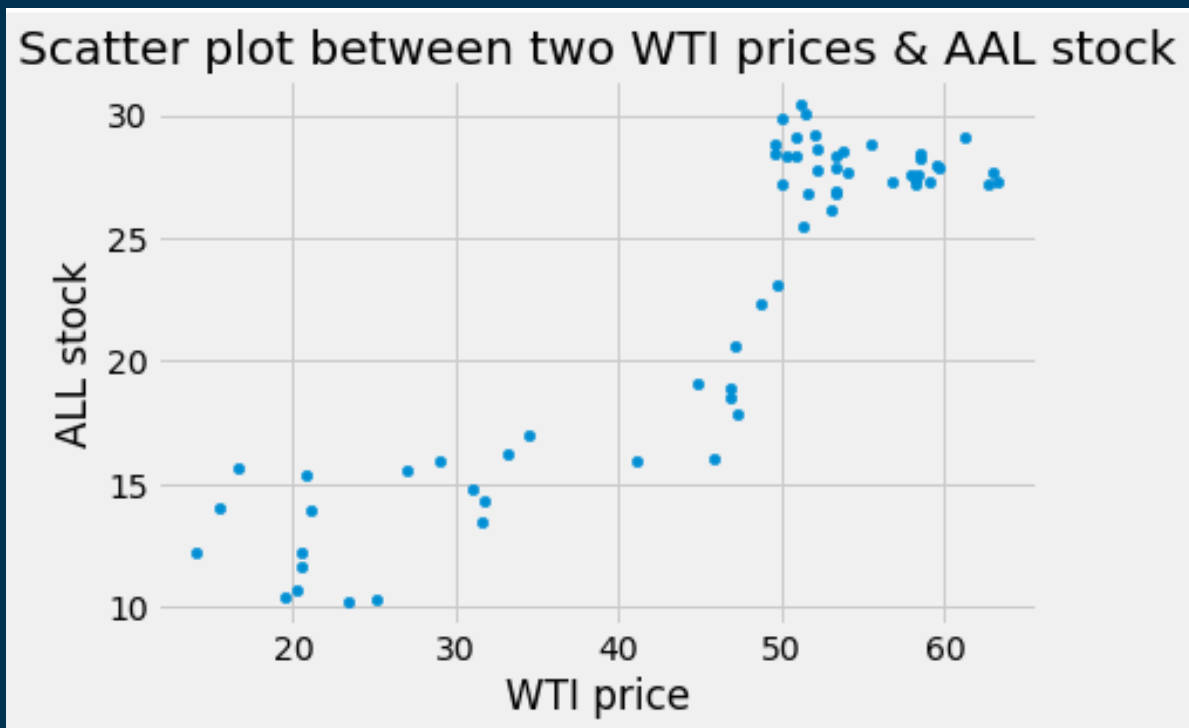
COVID-19 effects on the economy



עבור הנפט אנו רואים שרוב הערכים נמצאים 30 ל 55 דולר
כאשר הממוצע לאורך זמן הינו 50
כאשר יש נפילה דרמטית עד לערכים מינימליים של 15
עבור מחיר המנייה של חברת התעופה אנו רואים
שהממוצע לאורך הזמן היה 30 ורוב הזמן זה נשאר כך
אך בעקבות משבר הקורונה נהיו ערכים מינימליים
שרחוקים מאוד מהטווח- הנקודות

גרף קופסה המתאר את מניית הנפט ואמריקן איירליינס בדולר

COVID-19 effects on the economy



בדיקה לגבי אופן השינוי של המחירים בעקבות משבר הקורונה
נראה שהייתה יציבות לפני שהקורונה החלה, הגוש הימני העליון
עם הזמן בחודשים האחרונים יש קריסה כללית בביקוש
לנפט- רוב הטיסות בעולם מושבתות ולכן גם אין רווח
מנסיגת מחיר הנפט



Checking correlations with a Linear Regression model

```
test_results.rootMeanSquaredError
```

```
4.368290850970048
```

```
test_results.r2
```

```
0.9227018185854577
```

```
final_data.describe().show()
```

```
+-----+-----+
|summary|      Price|
+-----+-----+
|  count|         63|
|   mean|  44.93746031746033|
| stddev| 14.339810810855973|
|    min|         14.1|
|    max|         63.27|
+-----+-----+
```

כאן השתמשנו במודל של רגרסיה לינראית בשביל לנסות לתאר את הקשר בין מחירי מניית הנפט לבין כמות הדבקים והמתים במחלת הקורונה.

מצאנו לפי הנתונים כי יש קורלציה שלילית של -0.8 בין מחיר הנפט לבין כמות החולים והמתים בקורונה וכן בין מצאנו כי גם קיים קורלציה חיובית גבוהה של 0.9 בין מחיר הנפט לבין מניית אמריקן איירליינס.

המודל הצליח לתאר באופן יחסית מוצלח את הקשר בין מחירי מניות לבין כמות נדבקים ומתים מקורונה.

תודה רבה