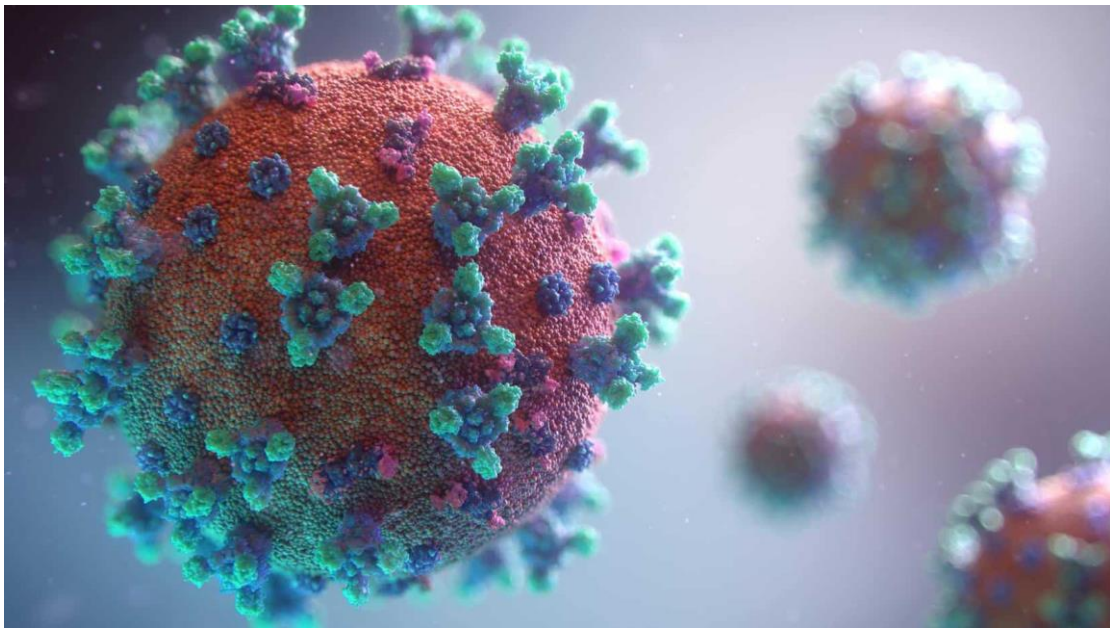


פרויקט סיום ויזואליזציה



גלי קולני 208156463
יהל חותם 213508096
רותם קלר 314949496

תוכן עניינים :

הקדמה.....	עמוד 3
תיאור נתוני העבודה.....	עמוד 4
רעיונות ותכנונים.....	עמוד 5-6
הצגת הפתרון שלנו.....	עמוד 7-8
יתרונות+חסרונות.....	עמוד 9-10
הסבר על היישום.....	עמוד 11-12
ויזואליזציות.....	עמוד 13-26

1.הקדמה :

לינק לפרויקט :

https://public.tableau.com/views/Project_16858882714730/Covid19Story?language=en-US&:display_count=n&:origin=viz_share_link

הצגת הנושא: נתוני קורונה איזורים סטטיסטיים בישראל-

[README - Government Data](#) - נתוני קורונה איזורים סטטיסטיים - COVID-19 מאגר
[covid-19 by area - Government Data](#) - נתוני קורונה איזורים סטטיסטיים - COVID-19 מאגר

מגפת הקורונה הייתה מגפה עולמית של מחלת נגיף הקורונה. במגפה נדבקו מעל מיליארד בני אדם, מתוכם 680 מיליון מאומתים, ומתו לפחות 6.9 מיליון, אם כי לפי הערכות שונות מספר הנדבקים והמתים גבוה בהרבה. המגפה התחילה בשנת 2020. הנושא שבחרנו לפרויקט הוא נתוני קורונה לפי אזורים גיאוגרפים, נתוני מאומתים ומחלימים לפי אזורים סטטיסטיים של הלמ"ס, מסודרים לפי תאריך.

בחירת הנושא :

בחרנו נושא זה מכיוון שנושא זה הוא נושא מרכזי בחיינו בשנים האחרונות. הנושא מעניין אותנו, ורצינו לחקור את הנושא בעצמנו מנקודת מבט שכל אחד מחברי הקבוצה פחות התעניין בה בחיי היום יום שלו בזמן המגיפה ובכלל. בפרויקט שלנו, בחרנו להתמקד באספקט הגיאוגרפי. בנוסף לכך, בחרנו גם את הנושא כי לדעתנו אפשר לייצר לנושא זה ויזואליזציות יפות, מעניינות, משמעותיות וברורות.

מגפת הקורונה הינה מגיפה עולמית אשר משפיעה על כל העולם, וכמובן שגם על אזרחי ישראל. יש דרכים בהן אפשר לנסות לצמצם את המגיפה, ללמוד דפוסים ותבניות מסוימות שיוכלו לעזור לאנשים ללמוד כיצד לנהוג על מנת להפחית את הסיכוי להידבקות והפרויקט שלנו עוסק בנושא זה. לכן, נושא זה חשוב מאוד לדעתנו.

משתמשים פוטנציאליים :

המשתמשים הפוטנציאליים בפרויקט שלנו יכולים להיות מספר קבוצות שונות. ראשית, המשתמשים יכולים להיות כלל האוכלוסייה בישראל. אזרחי המדינה יכולים להיעזר בפרויקט זה על מנת לדעת באילו אזורים גיאוגרפים יש יותר מאומתים לקורונה באופן כללי, באיזה אזורים המאומתים חולים ברמה קשה יותר, מספר המחוסנים באותו אזור, מספר המחלימים וכו'. למשל, אם אזרחים רוצים לעבור דירה, והם רוצים לחפש אזור בו התחלואה נמוכה יחסית, פרויקט זה יכול לעזור להם להבין איזה אזורים יכולים להתאים להם בצורה קלה וברורה.

שנית, משתמשים פוטנציאליים נוספים יכולים להיות מומחי קורונה, חוקרי קורונה, אשר יכולים בעזרת הפרויקט לגלות מגמות מעניינות ודפוסים מעניינים בצורה קלה.

נוסף לכך, בתי החולים יוכלו להיעזר בפרויקט זה. למשל, אם לשני בתי חולים יש מספר רב של מאושפדים, הם יוכלו לדבר ביניהם ולהתייעץ לגבי דרכי התנהלות. זאת ועוד, בתי חולים שאם בהם הרבה מאושפדים, יוכלו לעזור לבתי חולים עם מספר רב של מאושפדים למשל. לסיכום, פרויקט זה פונה למספר קהלי יעד שונים. לכל קהל יעד יש מטרות שונות ולכן חלקים שונים של הפרויקט יתאימו למשתמשים שונים.

2. תיאור נתוני העבודה:

את המידע לפרויקט, לקחנו מהאתר למאגרי המידע הממשלתיים,

[Government Data - מאגרי מידע - COVID-19 מאגר](#)

רקע לנתונים:

טבלת אזורים סטטיסטיים

קוד יישוב	שם הישוב	אזור סטטיסטי	תאריך	מספר מקרים מאומתים מצטבר עד התאריך הנקוב (כולל)	האם ישנה תוספת של חולים חדשים בתאריך?	מספר מחלימים מצטבר עד התאריך הנקוב (כולל)	האם ישנה תוספת של מחלימים חדשים בתאריך?	מספר מאושפדים מצטבר עד התאריך הנקוב (כולל)	האם ישנה תוספת של מאושפדים חדשים בתאריך?	מספר נפטרים מצטבר עד התאריך הנקוב (כולל)	האם ישנה תוספת של נפטרים חדשים בתאריך?	מספר בדיקות אבחנה מצטבר עד התאריך הנקוב (כולל)
קוד 2-4 ספרות	שם הישוב (כפי הישוע בטבלאות שמופיע בה'ס)	קוד 4 ספרות	DD/MM/YYYY	מספר רצף	בוליאני -	מספר רצף	בוליאני -	מספר רצף	בוליאני -	מספר רצף	בוליאני -	מספר רצף
רק של יישובים גדולים מ-2000 תושבים על פי טבלת ה'ס	שם הישוב (כפי הישוע בטבלאות שמופיע בה'ס)	קוד 4 ספרות	מוצג החל למרץ 11-2000	מספר רצף	בוליאני -	מספר רצף	בוליאני -	מספר רצף	בוליאני -	מספר רצף	בוליאני -	מספר רצף

מספר מצטבר של המתחסים במנה שנית עד התאריך הנקוב	מספר מצטבר של המתחסים במנה שנית עד התאריך הנקוב	מספר מצטבר של המתחסים במנה שנית עד התאריך הנקוב
15 < או מספר רצף החל מה- 30.7.2021	15 < או מספר רצף החל מה- 10.1.2021	15 < או מספר רצף החל מה- 20.12.2020

בטבלה זו ניתן למצוא ברמה יומית וגיאוגרפית נתונים לגבי הכמות המצטברת (עד לתאריך הרלוונטי) של מאומתים, מחלימים, מאושפדים, נפטרים ובדיקות אבחון למאומתים. כמו כן, ניתן למצוא ברמה יומית את המספר המצטבר של המחוונים במנה אחת ושתי מנות. המיקום הגיאוגרפי נקבע על פי כתובת המגורים המעודכנת של האזרח/ית. היחידה הגיאוגרפית המוצגת היא קוד ישוב ואזור סטטיסטי. האזורים הסטטיסטיים הם יחידות גאוגרפיות קטנות והומוגניות ככל האפשר, ובאמצעותם ניתן לשקף את המאפיינים הייחודיים של אזורים בתוך יישוב שבו יותר מ-10,000 תושבים. אזור סטטיסטי מונה בדרך כלל 3,000-5,000 תושבים. יישובים מתחת ל-10,000 תושבים מיוצגים באזור סטטיסטי אחד או רק על ידי קוד ישוב. לשם שמירה על פרטיות האזרחים, בתאים (חיתוכים של אזור סטטיסטי ותאריך) בהם המספר המצטבר של המאומתים/מחלימים/מאושפדים/נפטרים/בדיקות קטן מ-15 מצוין <15. בקובץ נתונים המקורי יש 1,942,906 רשומות.

3. רעיונות ותכנונים:

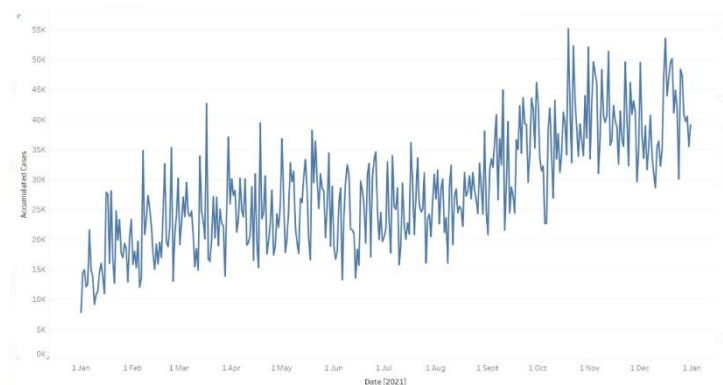
- במחשבה ראשונית, כאשר התחלנו לעבוד על הפרויקט, לא ידענו איך להתמודד עם כמעט 2 מיליון רשומות מכיוון שזהו מידע מאוד גדול וכמובן שלא הצלחנו לראות את קובץ האקסל המלא.
- תחילה, חשבנו לצמצם את המידע ולסנן אותו. עלו לנו מספר רעיונות לסינון המידע על מנת שיהיה יחסית מייצג.
- תחילה, לקחנו רק 20,000 רשומות משנת 2021 כלומר, מקרים משנת 2021. סיננו את הנתונים על ידי השאילתה :

https://data.gov.il/api/3/action/datastore_search?resource_id=d07c0771-01a8-43b2-96cc-c6154e7fa9bd&q=2021&limit=20000

כלומר, שאילתה זו מסננת לפי שנת 2021 ועם גבול של 20,000 תוצאות.

מידע זה לא היה מייצג מספיק, זה בערך אחוז ממספר הרשומות הכללי ולכן ברור שלא מייצג.

כאשר ניסנו לייצר גרף המראה את כמות המקרים המצטברים לפי חודשים, הגרף יצא לא הגיוני ומאוד רועש :



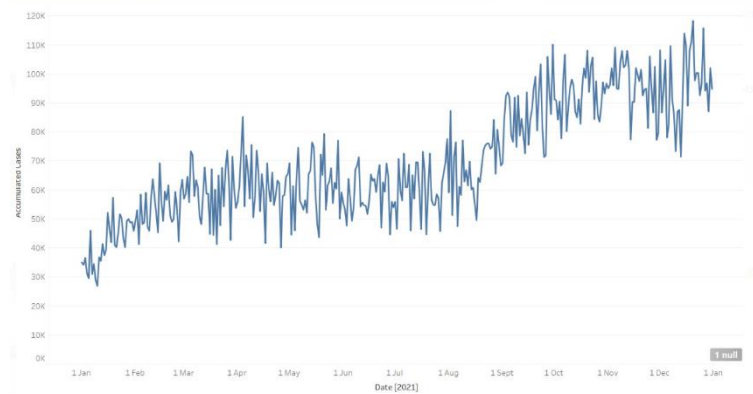
ניתן לראות שקיימות בגרף ירידות, דבר שהוא לא הגיוני כי זה מקרים מצטברים.

- לאחר מכן, לקחנו 50,000 רשומות משנת 2021, אך גם זה לא היה מייצג מספיק. סיננו את הנתונים על ידי השאילתה :

• https://data.gov.il/api/3/action/datastore_search?resource_id=d07c0771-01a8-43b2-96cc-c6154e7fa9bd&q=2021&limit=50000

כלומר, שאילתה זו מסננת לפי שנת 2021 ועם גבול של 50,000 תוצאות.

ניתן לראות שגרף זה פחות רועש מהקודם, אבל עדיין מאוד רועש. בנוסף, ניתן לראות שגם בגרף זה קיימות ירידות, דבר שהוא לא הגיוני.



- ניתן לראות משני הגרפים של המקרים המצטברים לפי חודשים שזה לא הגיוני שקיימות ירידות, הגרפים אמורים להיות גרפים עולים. דבר זה קרה בגלל שלא לקחנו את כל המידע והיו חסרות המון נקודות ולכן זה לא יצא גרף עולה.
- בסופו של דבר, הבנו שלסנן את הנתונים בצורה שלא פוגעת במידע זה מאוד קשה, ועל מנת לשמור על כל המידע כך שיהיה מייצג ונכון החלטנו להשתמש בכל 2 מיליון הרשומות.

- קושי נוסף שהתמודדנו איתו הוא שרצינו להציג את מפת ארץ ישראל עם כל הערים והישובים הקיימים בדאטה על המפה. תחילה, ניסינו להציג את הערים והישובים על המפה לפי שם העיר אך היו ישובים שטאבלו לא הכיר.
- **התמודדות:** התמודדנו עם קושי זה על ידי כך שחיפשנו את הקורדינטות של כל הערים והישובים בדאטה שלנו, וכך הכנסנו את המידע לטאבלו. השתמשנו בקוד שכתבנו על מנת להמיר את הקורדינטות לפורמט שטאבלו מכיר. הסבר מפורט נוסף על זה יהיה בחלק של ה-preprocessing. תהליך זה היה מאוד ארוך בפרויקט שלנו.

4.הצגת הפתרון שלנו :

- בפתרון שלנו בחרנו להתמקד בשני אספקטים עיקריים:
האספקט הראשון הוא המיקום הגיאוגרפי והאספקט השני הוא הזמן.
שאלות המחקר שהנחנו אותנו הן:
 - 1. האם אזור המגורים משפיע על כמות החולים בקורונה? כלומר, השפעה של האזור הגיאוגרפי על כמות החולים בקורונה.
 - 2. האם אזור המגורים משפיע על חומרת וירוס הקורונה? (יותר מאושפזים, יותר מקרי מוות וכו').
 - 3. כיצד מספר החולים בקורונה משתנה לאורך הזמן?
 - 4. כיצד מספר המחלימים והחיסונים משתנה לאורך זמן?
- בפרויקט זה, בחרנו לממש את הוויזואליזציה שלנו בעזרת ציר זמן ומפה אשר שולטים על שאר האלמנטים .
כמו כן, בחרנו במפה כי היא מאפשרת השוואה בין הערים/הישובים השונים, היא קלה להבנה ולשימוש.

• הויזואליזציות שלנו :

1. **מפת ארץ ישראל : Area Vs Cases** – ויזואליזציה זו מראה את השפעת האזור הגיאוגרפי על מספר מקרי הקורונה. סימנו על מפת ארץ ישראל את כל הישובים/הערים שהיו לנו בדאטה. צבענו כל אחת מהנקודות על המפה בצבע בגווי אדום לפי מספר החולים בקורונה בישוב זה. השתמשנו במפת אדומים כך שכל שהנקודה אדומה יותר , מספר החולים בקורונה בעיר זו גבוה יותר.
2. **מפת ארץ ישראל: Area Vs Deaths/Hospitalization** -ויזואליזציה זו מראה את השפעת האזור הגיאוגרפי על חומרת וירוס הקורונה. בחרנו להראות את חומר וירוס הקורונה על ידי שני פרמטרים : האחד הוא מספר המאושפזים באותו אזור והשני הוא מספר מקרי המוות באותו אזור. סימנו על מפת ארץ ישראל את כל הישובים/הערים שהיו לנו בדאטה. בויז זה מיפינו את צבע הנקודות במפה למספר מקרי המוות ואת גודל הנקודה למספר המאושפזים באותו ישוב. מפת הצבעים שבחרנו היא אפור-שחור. ככל שהנקודה שחורה יותר, כהה יותר, מספר מקרי המוות בישוב זה גדול יותר . בנוסף, ככל שהנקודה גדולה יותר , מספר המאושפזים בעיר זו גדול יותר.
3. **Cases Vs Time :Line chart** -גרף זה מראה את השינוי של מספר חולי הקורונה לאורך זמן.
ציר ה-x ממופה לחודשים פברואר, יוני ואוקטובר בכל אחת משנים 2020-2023, וציר ה-y ממופה למספר חולי הקורונה במאות אלפים. בגרף זה בחרנו במפת צבעים של אדום. ככל האדום כהה יותר, כך מספר החולים גבוה יותר. גרף זה הוא גרף פשוט, הממחיש בצורה ברורה עלייה בסך כל חולי הקורונה. ניתן ללמוד מגרף זה באילו חודשים/תקופות זמן היו עליות חדות יותר במספר החולים ובאילו מתונות יותר.

4. **Deaths Vs Time: Line chart** - גרף זה מראה את השינוי של מספר המתים מקורונה לאורך זמן.
ציר ה-x ממופה לחודשים פברואר, יוני ואוקטובר בכל אחת משנים 2020-2023, וציר ה-y ממופה למספר המתים. בגרף זה בחרנו במפת צבעים של אפור-שחור. ככל שהצבע כהה יותר, כך מספר המתים גבוה יותר. גרף זה הוא גרף פשוט, הממחיש בצורה ברורה עלייה בסך כל המתים מקורונה. ניתן ללמוד מגרף זה באילו חודשים/תקופות זמן היו עליות חדות יותר במספר המתים ובאילו מתונות יותר.
5. **Recoveries \ Vaccinations Vs Time: Line chart** - גרף זה מראה את השינוי במספר המחלימים ובמספר המתחסנים לאורך זמן. בגרף זה ישנם 4 קווים: שלושת הקווים בגווי הכחול הם מספר המתחסנים בחיסון הראשון, השני והשלישי. ניתן להבדיל בין שלושת החיסונים גם לפי הצבע והמקרא ובנוסף גם לפי הזמן, ברור שהחיסון הראשון היה בתאריך הכי מוקדם, השני אחריו ואז השלישי. הקו הרביעי, הבולט מבין הארבעה, צבוע בצבע ירוק והוא מסמל על מספר המחלימים הכולל. ציר ה-x ממופה לתאריכים וציר ה-y ממופה למספר המחלימים או המתחסנים.
6. **Top Cities :Bar chart** - גרף זה מראה את k הערים עם הכי הרבה מקרי קורונה. בגרף זה בחרנו במפת צבעים של אדום, ככל שהעמודה אדומה יותר, כהה יותר, מספר החולים גדול יותר. ציר ה-x מראה את שמות הערים וציר ה-y מראה את מספר החולים. אנחנו בחרנו להציג את $K=3$ הערים עם הכי הרבה חולים אבל בצד ימין למעלה ניתן לבחור את הפרמטר K לפי הצורך. אנחנו בחרנו ב-K קטן מכיוון שבצורה זו הגרף מאוד קריא וברור.
7. **Covid By Time : Dashboard** - גרף זה מחולק לשלושה חלקים והוא מראה מה הקשר בין מספר החולים לפי האספקט של אזור גיאוגרפי לבין האספקט של הזמן. כלומר, dashboard זה מקשר בין שני האספקטים המרכזיים שלנו, אזור גיאוגרפי וזמן. גרף זה מורכב משלושת הגרפים שהסברנו עליהם למעלה, גרפים 1,3,6. ב-dashboards זה, ניתן לבחור נקודה מסוימת על ציר הזמן ולראות כיצד מספר חולי הקורונה משתנה מבחינת האזורים הגיאוגרפיים ומבחינת הערים עם כמות החולים הכי גדולה. כלומר ניתן לסנן, לפי נקודת זמן מסוימת.
8. **Covid By City : Dashboard** - גרף זה מחולק לשלושה חלקים והוא מראה מה הקשר בין מספר החולים לפי האספקט של אזור גיאוגרפי, לפי עיר ולפי זמן. כלומר, dashboard זה מקשר בין שני האספקטים המרכזיים שלנו, אזור גיאוגרפי וזמן. גרף זה מורכב משלושת הגרפים שהסברנו עליהם למעלה, גרפים 1,3,5. ב-dashboards זה, ניתן לבחור נקודה מסוימת על המפה, כלומר לבחור עיר מסוימת ולראות כיצד מספר חולי הקורונה, מספר המתחסנים (בשלוש המנות) ומספרים המחלימים משתנה בעיר זו לאורך הזמן. כלומר, dashboard זה נותן לנו את היכולות להתמקד בעיר מסוימת בצורה ממוקדת יותר ולחקור אותה. כלומר, להסתכל על המיקרו, על עיר ספציפית, ולא על התמונה/המצב הכללי.
9. **Covid19 Story : Story** - סיפור זה מסכם ומאחד לנו את כל הפרויקט. בסיפור זה ניתן לראות את הגרפים השונים שעבדנו עליהם במהלך הפרויקט בצורה מסודרת ומאורגנת, מרוכזים יחד עם הסברים נוספים.
- לכל הויזואליזציות יש tool tip .
 - **חשוב לשים לב** שההתמקדות של הויזואליזציות של המפות וה-bar chart המראה Top cities היא להראות את התפשטות המחלה ככלל על המדינה ולא התפשטות

- המחלה כפרט בכל עיר. לכן, בחרנו להשתמש בנתונים המלאים ולא להמיר לאחוזים של כל עיר על מנת להראות את השינוי הכללי על רחבי המדינה ולהראות את המוקדים של אוכלוסיות החולים לפי נקודות בזמן.
- נוסף לכך,** כך אנו שומרים על אחידות עם הדרכים של ה-line chart בהם לא ניתן להמיר לאחוזים.

יתרונות :

- הוויזואליזציות פשוטות, מובנות וברורות ולכן נגישות לכל סוגי המשתמשים.
- הצגה דינמית של הנתונים בעזרת ציר זמן, מאפשרת לראות נתונים מנקודות זמן שונות בתקופת הקורונה.
- ניתן לראות מגמות של תחלואה בקרב ערים/ישובים שונים.
- ניתן לראות מגמות של רמת החומרה של הווירוס בקרב ערים/ישובים שונים.
- ניתן להבין בפשטות באילו ערים/ישובים יש רוב של חולי קורונה(בעזרת המפה או ה-bar chart)/מאושפצים/מתים(באמצעות המפה וה-dashboard).
- ניתן להשוות בקלות בין ערים/ישובים באמצעות המפות ובאמצעות ה-dashboard.
- ניתן ללמוד דפוסים שונים על נקודות זמן בהן הייתה עלייה משמעותית או מתונה במספר החולים.
- בחירת צבעים בצורה מתאימה ופשוטה להבנה. למשל, ככל שיותר אדום יותר חולים, ככל שיותר שחור יותר מתים. בנוסף, כאשר רצינו להדגיש את מספר המחלימים לעומת מספר המתחסנים צבענו את הגרף של המחלימים בירוק, צבע הבולט בגרף זה.

חסרונות :

- משתמש שאין לו ידע נרחב בגיאוגרפיה יצטרך לחפש ערים/ישובים ידנית במפה.
- יש מספר גדול של ערים/ישובים, 260, ולכן לא ניתן להציג את כולם בצורה ברורה ב-bar chart.

חסרונות ובעיות עם הדטה(כללי):

- כשהתחלנו לעבוד על הפרויקט, גילינו כי יש הרבה דברים בעייתיים בדאטה שלנו שנצטרך לסדר.
- הדאטה מאוד גדול, כמעט 2 מיליון רשומות.
- בהתחלה, החלטנו לסנן את הדאטה בצורה חכמה, כזו שלא פוגעת במדגם.
- החלטנו לסנן את הדאטה לפי התאריך. בחרנו 50,000 רשומות משנת 2021.
- המדגם לא היה מספיק מייצג, ולכן בסופו של דבר בחרנו לעבוד עם כל ה-2 מיליון רשומות.
- אחד הוויזים שרצינו לעשות, הוא מפת הארץ עליה מסומנים כל הישובים והערים בדאטה לפי מספר מקרי הקורונה. שמות הערים בדאטה כתוב בעברית. הטאבלו

היא לא תוכנה התומכת בעברית, ולכן היינו צריכים לתרגם את כל שמות הערים מעברית לאנלית.

- בעיה נוספת שהייתה לנו היא שהיו המון ערים/ישובים, שגם שאחרי שתרגמנו את השם שלהם לאנגלית, טבלו לא הכירה אותם ולכן הם לא סומנו על המפה כמו שרצינו.
פתרנו את זה על ידי כך שאספנו את כל הקורדינטות של כל אחד מהישובים/ערים, כתבנו קוד פייתון המפריד את הקורדינטות לפי הפורמט שהטאבלו מקבל.
הוספנו שתי עמודות בדטה של קורדינטות של כל עיר או ישוב.
- נציין שוב, שהקורדינטות כמו שמצאנו אותן באינטרנט לא רשומות באותו פורמט שהטאבלו מקבל. לכן, היינו צריכים לכתוב קוד פיתון אפשר ממיר לצורה בטבלו מקבל.
- **כל אלה דרשו מאיתנו המון עבודה.** בדאטה שלנו יש 260 ערים/ישובים שונים.

5. הסבר על היישום:

עיבוד הנתונים:

- לפני שהתחלנו לעבוד על הוזים השונים, היינו צריכים לעבד את הנתונים.

:Pre-processing

- בחלק מהוזים בפרויקט שלנו, התמקדנו בהשפעה של האזור הגיאוגרפי על כמו חולי הקורונה, חומר המחלה וכו'. לכן, היינו צריכים להכין את הדאטה בצורה טובה במטרה שנוכל להראות היבטים הקשורים למיקום הגיאוגרפי.
- תחילה, הוספנו לדאטה שלנו עמודה של שם המדינה, ישראל לכל הרשומות מכיוון שהדאטה שלנו הוא מישראל.
- שנית, שמות הערים והישובים בדאטה שלנו היו בעברית. לכן, היינו צריכים לתרגם שמות של 260 ערים וישובים מעברית לאנגלית על מנת שהטאבלו יוכל לקלוט, להבין שמות אלה.
- בנוסף לכך, היו ישובים וערים שהטאבלו לא הכיר את השמות שלהם, לכן הוספנו עוד שתי עמודות בדאטה שלנו, של הקורדינטות של כל ישוב.
- תהליך זה היה ארוך ממספר סיבות: ראשית, בדאטה שלנו היו 260 ערים/ישובים שונים.
- שנית, גם אחרי שמצאנו את הקורדינטות באינטרנט, הם לא היו בפורמט שהטאבלו מקבל והיינו צריכים להפוך את זה פורמט שטבלו מקבל. כתבנו קוד פייתון העושה זאת.

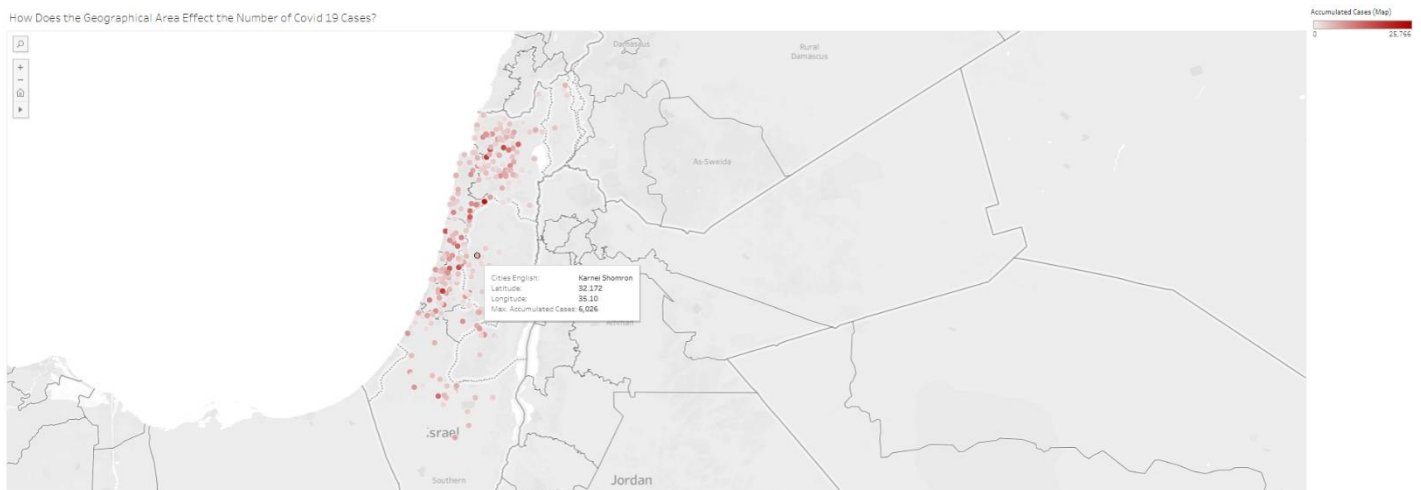
הסבר למשתנים בטבלה:

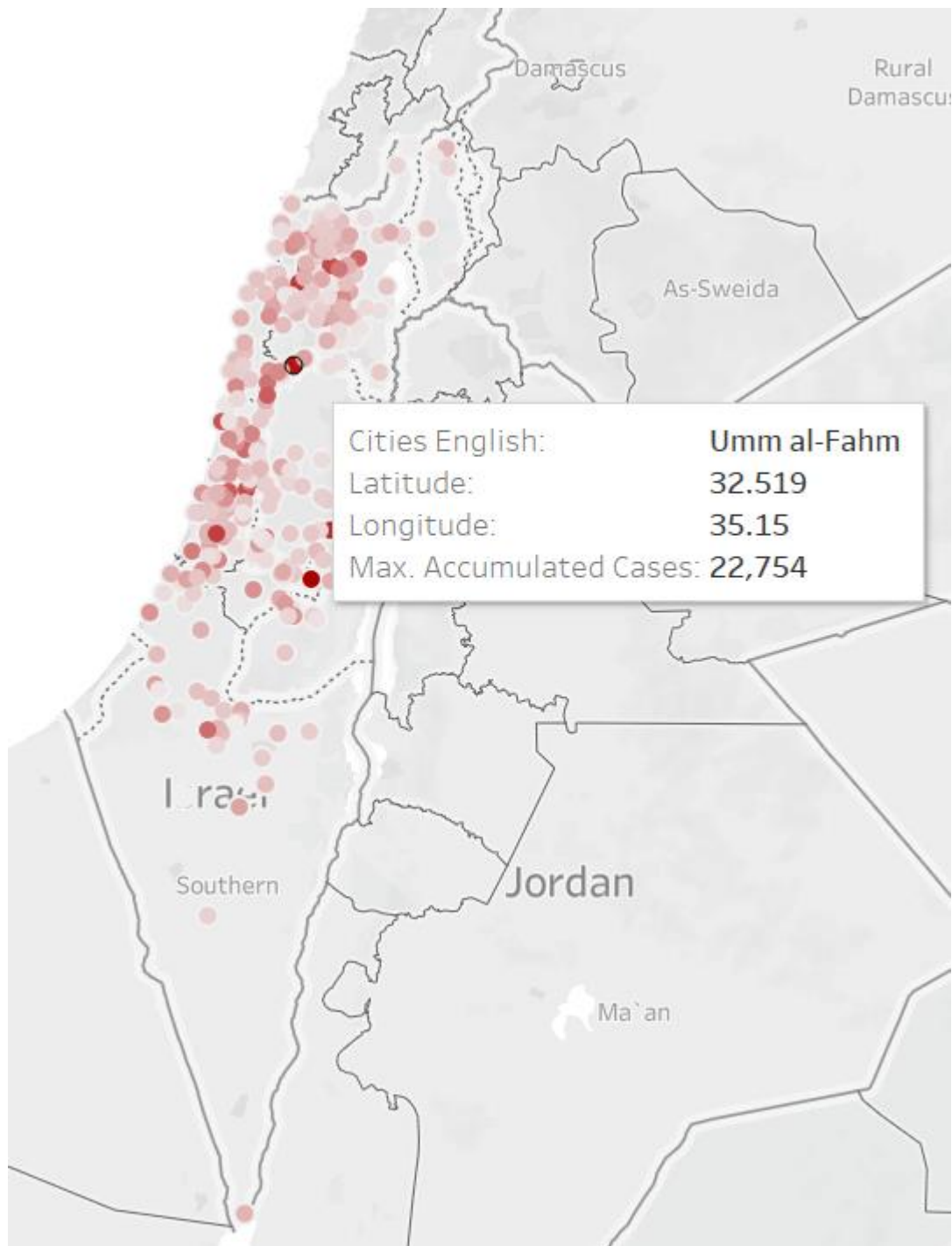
1. Town_code: קוד הישוב כפי שמוגדר בטבלאות [הלמ"ס](#).
2. Town: שם הישוב.
3. Agas_code: קוד האזור הסטטיסטי המיוצג על ידי 4 ספרות.
4. date: במאומתים – מייצג את התאריך בו הבדיקה החיובית הראשונה התקבלה במעבדה; במחלימים – מייצג את תאריך ההחלמה (על פי המדיניות הנהוגה בתאריך), בנפטרים את תאריך התמותה, במאושפזים את תאריך תחילת האשפוז, בבדיקות את תאריך קבלת התוצאה במעבדה.
5. Accumulated_cases: מספר המאומתים המצטבר לאזור סטטיסטי עד לתאריך הרלוונטי (כולל המחלימים והנפטרים).
6. new_cases_on_date: האם נוספו מאומתים חדשים באותו היום. בוליאני – TRUE/FALSE.
7. Accumulated_recovered: מספר המחלימים המצטבר לאזור סטטיסטי עד לתאריך הרלוונטי.
8. new_recovered_on_date: האם נוספו מחלימים חדשים באותו היום. בוליאני – TRUE/FALSE.
9. accumulated_hospitalized: מספר המאושפזים המצטבר לאזור סטטיסטי עד לתאריך הרלוונטי (כולל אלו שהשתחררו).
10. new_hospitalized_on_date: האם נוספו מאושפזים חדשים באותו היום. בוליאני – TRUE/FALSE.
11. accumulated_deaths: מספר הנפטרים המצטבר לאזור סטטיסטי עד לתאריך הרלוונטי.
12. new_deaths_on_date: האם נוספו נפטרים חדשים באותו היום. בוליאני – TRUE/FALSE.
13. accumulated_diagnostic_tests: מספר בדיקות מצטבר לאיתור מאומתים באזור הסטטיסטי עד לתאריך הרלוונטי. בדיקות איתור מאומתים, בשונה מבדיקות החלמה מוגדרות בקרב נבדקים אשר אין להם תשובה חיובית קודמת. לדוגמא: אם אדם עשה בדיקה באפריל ויצא שלילי ועשה בדיקה בספטמבר כי נחשף לחולה מאומת ושוב יצא שלילי, אז שתי הבדיקות הן בדיקות לאיתור מאומתים ושתי הבדיקות יספרו בסכום המצטבר בהתאם לתאריכים הרלוונטיים. לעומת זאת, אם אדם עשה בדיקה במאי, ויצא חיובי ואז עשה שוב בדיקה על מנת לבחון תהליך החלמה, אזי הבדיקה הראשונה היא בדיקת איתור מאומתים ונספרת לסכום המצטבר בתאריך הרלוונטי במאי, ואילו הבדיקה השנייה אינה בדיקת איתור, אלא בדיקת החלמה ולכן אינה נכללת בסכימה.
14. Accumulated_vaccination_first_dose: מספר מצטבר של המתחסנים במנה אחת לאזור סטטיסטי עד לתאריך הרלוונטי. הנתונים רלוונטים החל מה- 20.12.2020.
15. Accumulated_vaccination_second_dose: מספר מצטבר של המתחסנים במנה שניה לאזור סטטיסטי עד לתאריך הרלוונטי. הנתונים רלוונטים החל מה- 10.1.2021.
16. Accumulated_vaccination_third_dose: מספר מצטבר של המתחסנים במנה שלישית לאזור סטטיסטי עד לתאריך הרלוונטי. הנתונים רלוונטים החל מה- 30.7.2021.

6. ויזואליזציות:

בחלק זה של הפרויקט, נציג את הויזואליזציות השונות שיצרנו בפרויקט זה.

- ויזואליזציה זו מראה את השפעת האזור הגיאוגרפי על מספר מקרי הקורונה. במפה מסומנים כל היישובים שקיבלנו בדאטה. ניתן לקרב ולהרחיק. כאשר מקרבים ניתן לראות בצורה ברורה את הנקודה של כל אחת מהערים. המיפויים בויז זה : המשתנה הויזואלי של הצבע מקושר לכמות החולים. כלומר, בחרנו במפת צבעים אדומה . ככל שהנקודה של העיר היא אדומה יותר, כך מספר החולים בעיר רב יותר.



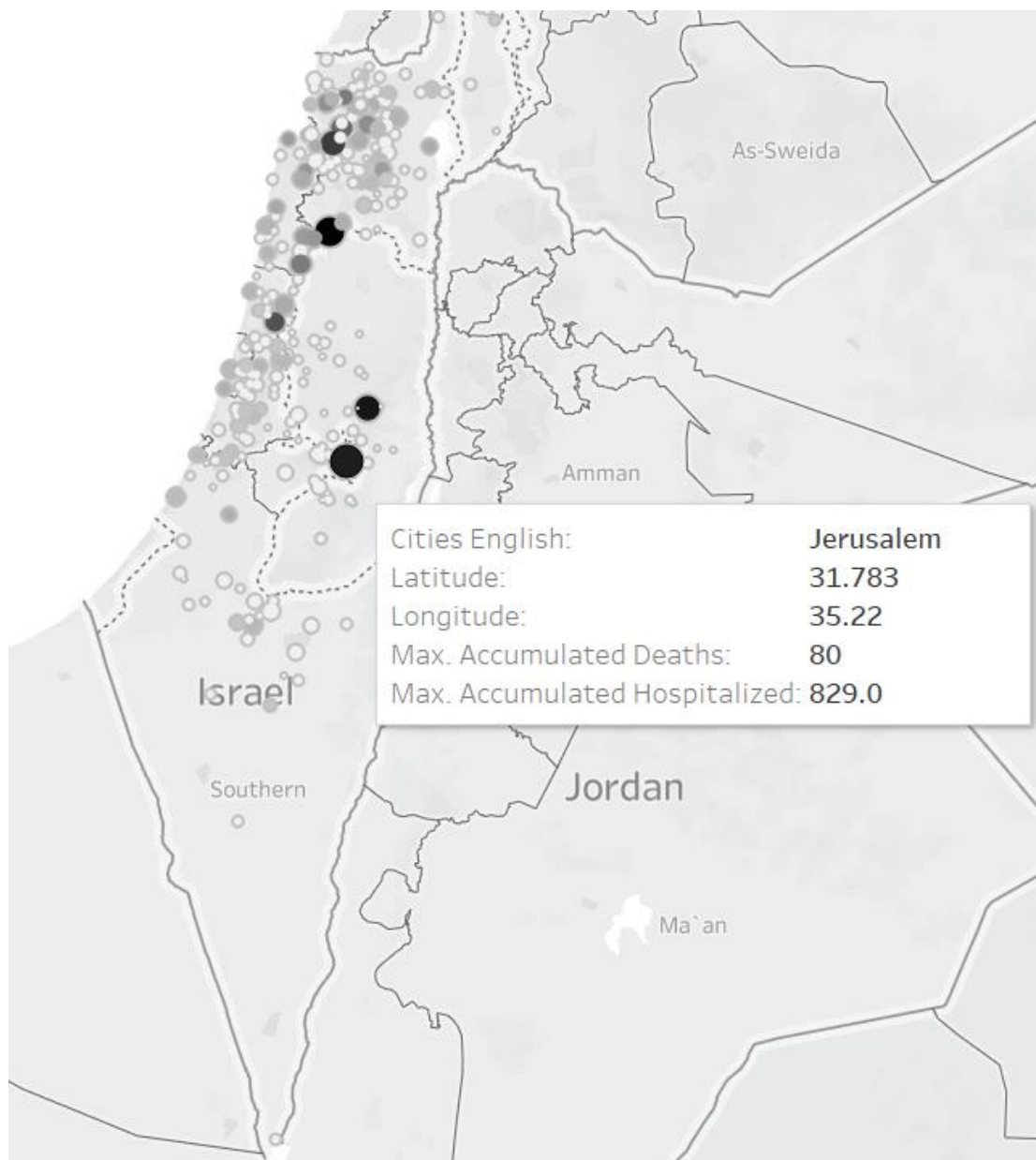


- כאשר עומדים על נקודה מסוימת בגרף זה, ניתן לראות מהו שם העיר, את הקורדינאטות של עיר זו ואת מספרי חולי הקורונה המצטבר בעיר.
 - האינטראקציה העיקרית בוויזואליזציה היא בין מיקום (עיר) לבין כמות המקרים המצטברת.
- יש קשר אינטראקטיבי בין וויזואליזציה זו לבין האחרות – ניתן לבחור רגע זמן מסוים בוויזואליזציית הגרף Cases vs Time, והמפה משתנה כך שמראה את כמות המקרים המצטברת עד לתאריך שנבחר בגרף.

- Task של וויזואליזציה זו: Explore Distribution
 - action של וויזואליזציה זו היא Explore – הפעולה שלוקחים בוויז היא מעבר על כל המפה, לא בהכרח עם התמקדות במיקום או אזור מסוים.
 - target של וויזואליזציה זו היא Distribution – באיזה אזורים או מקומות על המפה יש יותר או פחות חולים, מהי ההתפלגות הכללית של חולים על המפה. **חשוב לשים לב** שהמטרה העיקרית פה היא אינה למצוא את הנתונים על כל עיר עצמה, אלא את ההתפלגות הכללית על המפה, ולכן הנתונים נמצאים בתור מספר מקרים כללי ולא בתור אחוזים.

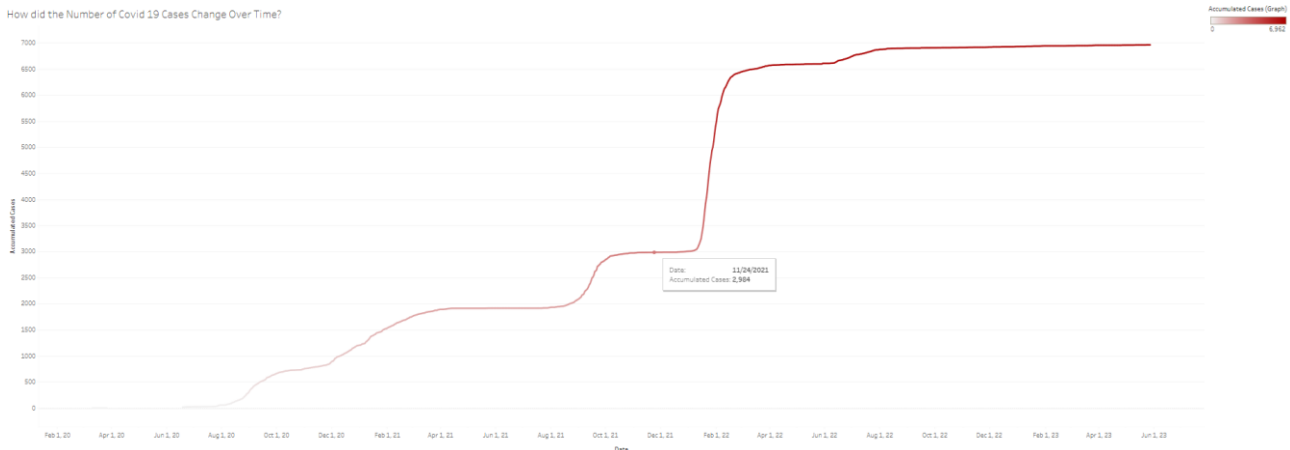
- וויזואליזציה זו מראה את השפעת האזור הגיאוגרפי על חומרת וירוס הקורונה. בחרנו להראות את חומר וירוס הקורונה על ידי שני פרמטרים : האחד הוא מספר המאושפדים באותו אזור והשני הוא מספר מקרי המוות באותו אזור. המיפויים בויז זה : החלטנו למפות את גודל הנקודה בה מסומנת כל עיר למספר המאושפדים בכל עיר ואת צבע הנקודה בה צבועה כל עיר למספר מקרי המוות בעיר זו. ככל שהנקודה שחורה יותר/כהה יותר, כל מקרי המוות בעיר זו רבים יותר.





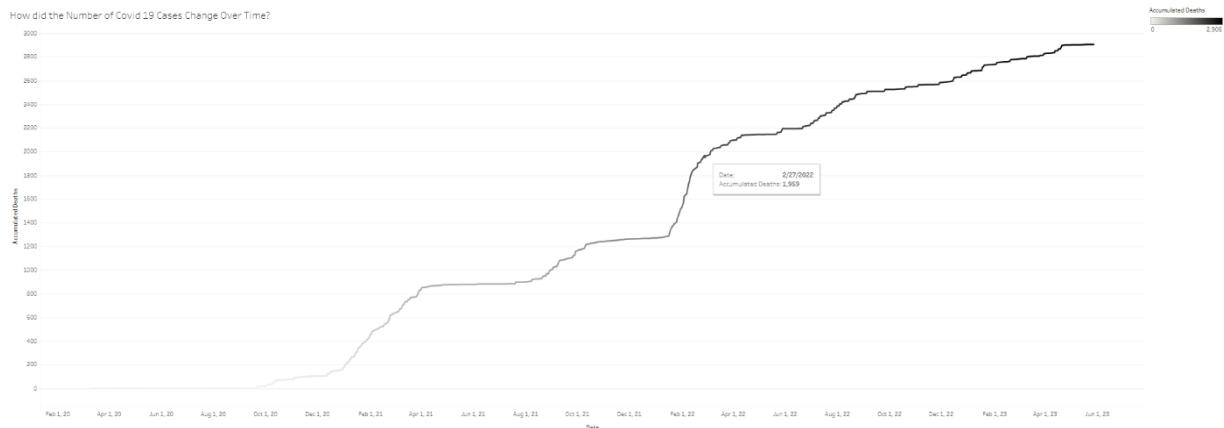
- כאשר עומדים על נקודה מסוימת בגרף זה, ניתן לראות מהו שם העיר, את הקורדינאטות של עיר זו, את מספרי מקרי המוות המצטבר ומספר המאושפזים המצטבר בעיר.
- האינטראקציה העיקרית בוויזואליזציה היא בין מיקום (עיר) לבין כמות המתים המצטברת וכמות המאושפזים המצטברת.
- Task של וויזואליזציה זו: Explore Distribution
 - action של וויזואליזציה זו היא Explore – הפעולה שלוקחים בוויז היא מעבר על כל המפה, לא בהכרח עם התמקדות במיקום או אזור מסוים.
 - target של וויזואליזציה זו היא Distribution – באיזה אזורים או מקומות על המפה יש יותר או פחות מקרי מוות \ אשפוז, מהי ההתפלגות הכללית של מקרי מוות \ אשפוז על המפה.

- גרף זה מראה את השינוי של מספר חולי הקורונה לאורך זמן. ציר ה-x ממופה לחודשים פברואר, יוני ואוקטובר בכל אחת משנים 2020-2023, וציר ה-y ממופה למספר חולי הקורונה במאות אלפים. מיפויים: בגרף זה מיפוינו צבע למספר חולי הקורונה באותו זמן. כלומר, ככל שהצבע יותר אדום, מספר חולי הקורונה רב יותר בתאריך זה.



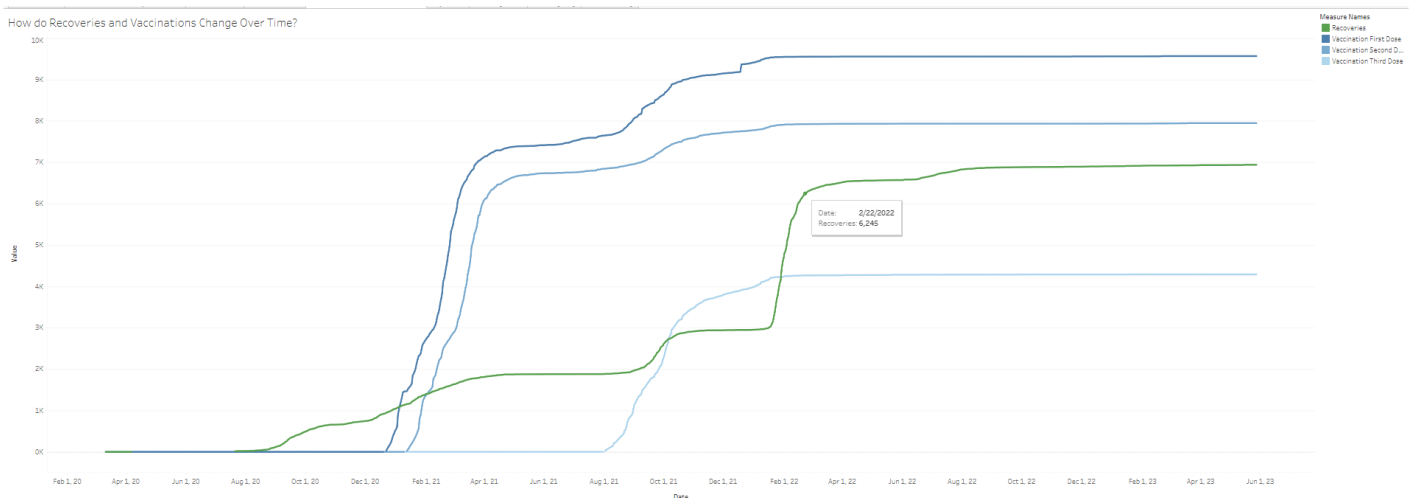
- ניתן לעבור על הגרף ולראות מהו מספר החולים המצטבר בתאריך מסוים.
- האינטראקציה העיקרית בוויזואליזציה היא זמן (תאריך מדויק) לבין כמות המקרים המצטברת
 - יש קשר אינטראקטיבי בין וויזואליזציה זו לבין האחרות – ניתן לבחור עיר מסוימת בוויזואליזציית המפה Area vs Cases, ובכך לסנן את הגרף, שייראה את כמות המקרים המצטברת עבור אותה העיר בלבד.
- Taskn של וויזואליזציה זו: Compare Trends
 - ההצגה של וויזואליזציה זו היא Compare Trends – בגלל האפשרות לסנן לפי עיר, הפעולה העיקרית של הוויזואליזציה היא השוואה בין עיר לבין כל עיר אחרת או לבין כל המדינה.
 - הtargetn של וויזואליזציה זו היא Trends – המטרה העיקרית של הוויזואליזציה היא השינוי של כמות המקרים המצטברת לאורך זמן, במילים אחרות trendn של כמות המקרים. הוויזואליזציה מראה את השינוי הכללי לאורך זמן, וגם נותנת למצוא נקודות מסוימות בזמן שבהן יש עליות משמעותיות.

- גרף זה מראה את השינוי של מספר המתים מווירוס הקורונה לאורך זמן. ציר ה-x ממופה לחודשים פברואר, יוני ואוקטובר בכל אחת משנים 2020-2023, וציר ה-y ממופה למספר המתים מקורונה. מיפויים: בגרף זה מיפוינו צבע למספר מקרי המוות מקורונה באותו זמן. כלומר, ככל שהצבע יותר כהה, יותר שחור, מספר מקרי המוות מקורונה רב יותר בתאריך זה.



- ניתן לעבור על הגרף ולראות מהו מספר המתים המצטבר בתאריך מסוים .
- האינטראקציה העיקרית בוויזואליזציה היא זמן (תאריך מדויק) לבין כמות מקרי המוות המצטברת
 - יש קשר אינטראקטיבי בין וויזואליזציה זו לבין האחרות – ניתן לבחור עיר מסוימת בוויזואליזציית המפה Area vs Cases, ובכך לסנן את הגרף, שייראה את כמות המקרים המצטברת עבור אותה העיר בלבד.
- Taskn של וויזואליזציה זו: Compare Trends
 - ההנחה של וויזואליזציה זו היא Compare – בגלל האפשרות לסנן לפי עיר, הפעולה העיקרית של הוויזואליזציה היא השוואה בין עיר לבין כל עיר אחרת או לבין כל המדינה.
 - הtargetn של וויזואליזציה זו היא Trends – המטרה העיקרית של הוויזואליזציה היא השינוי של כמות מקרי המוות המצטברת לאורך זמן, במילים אחרות trendn של מקרי המוות. הוויזואליזציה מראה את השינוי הכללי לאורך זמן, וגם נותנת למצוא נקודות מסוימות בזמן שבהן יש עליות משמעותיות.

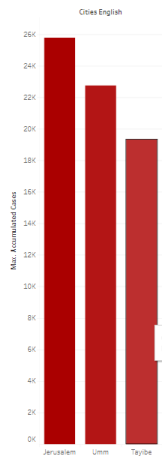
- גרף זה מראה את השינוי במספר המחלימים ובמספר המתחסנים לאורך זמן. בגרף זה ישנם 4 קווים: שלושת הקווים בגווני הכחול הם מספר המתחסנים בחיסון הראשון, השני והשלישי. ציר ה-x בגרף זה ממופה לתאריך לפי שנה וצריך ה-y בגרף זה ממופה לערך מספרי של מחלימים/מספר מתחסנים באחת ממנות החיסון. מיפויים: בחרנו למפות את הצבע למספר המתחסנים ובנוסף למספר המחלימים. את שלושת החיסונים עשינו בגוונים של כחול, ומה שחשוב היה לנו להראות מגרף זה הוא את מספר המחלימים, ולכן עשינו את גרף זה בצבע שונה, ירוק, הבולט ביחד לכל הגרף. כאשר עומדים על נקודה בגרף זה, ניתן לראות מהו התאריך ואת מספר המחלימים עד תאריך זה /מספר המתחסנים עד תאריך זה באחת מהמנות חיסון.



- האינטראקציה העיקרית בוויזואליזציה היא זמן (תאריך מדויק) לבין כמויות החיסונים מכל סוג וכמות המבריאים.
 - יש קשר אינטראקטיבי בין וויזואליזציה זו לבין האחרות – ניתן לבחור עיר מסוימת בוויזואליזציית המפה Area vs Cases, ובכך לסנן את הגרף, שייראה את הנתונים האלה עבור העיר הזאת בלבד.
- Task של וויזואליזציה זו: Compare Trends
 - ההנחה של וויזואליזציה זו היא Compare Trends – בגלל האפשרות לסנן לפי עיר, הפעולה העיקרית של הוויזואליזציה היא השוואה בין עיר לבין כל עיר אחרת או לבין כל המדינה.
 - הtarget של וויזואליזציה זו היא Trends – המטרה העיקרית של הוויזואליזציה היא השינוי של הנתונים לאורך זמן, במילים אחרות trend של כמות החיסונים וכמות המבריאים. הוויזואליזציה מראה את השינוי הכללי לאורך זמן, וגם נותנת למצוא נקודות מסוימות בזמן שבהן יש עליות משמעותיות בכל נתון.

- גרף זה מראה את k הערים עם הכי הרבה מקרי קורונה. בגרף זה ניתן לבחור את מספר הערים שרוצים להציג.
בגרף זה ציר ה- x מראה את שמות הערים וציר ה- y מראה את מספר החולים. מיפויים: בגרף זה בחרנו למפות את הצבע למספר החולים בעיר. ככל שמספר החולים בעיר גדול יותר, הצבע כהה יותר (אדום יותר).

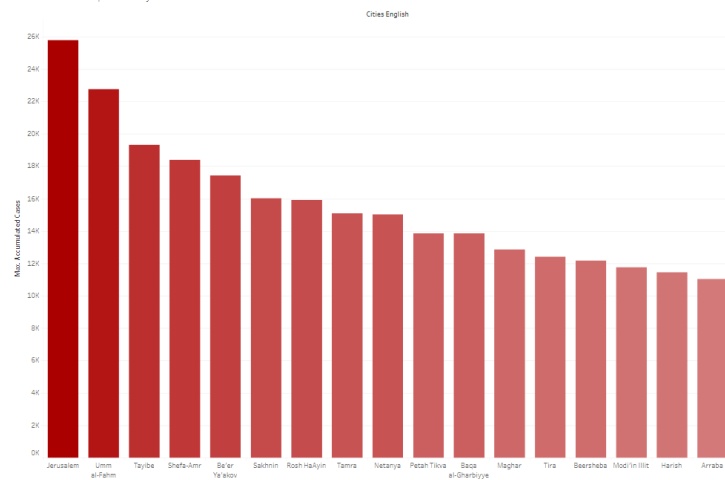
Which are the Top Cities by Covid Cases?



Top Cities Num
3
0

Accumulated Cases
0 25,700

Which are the Top Cities by Covid Cases?



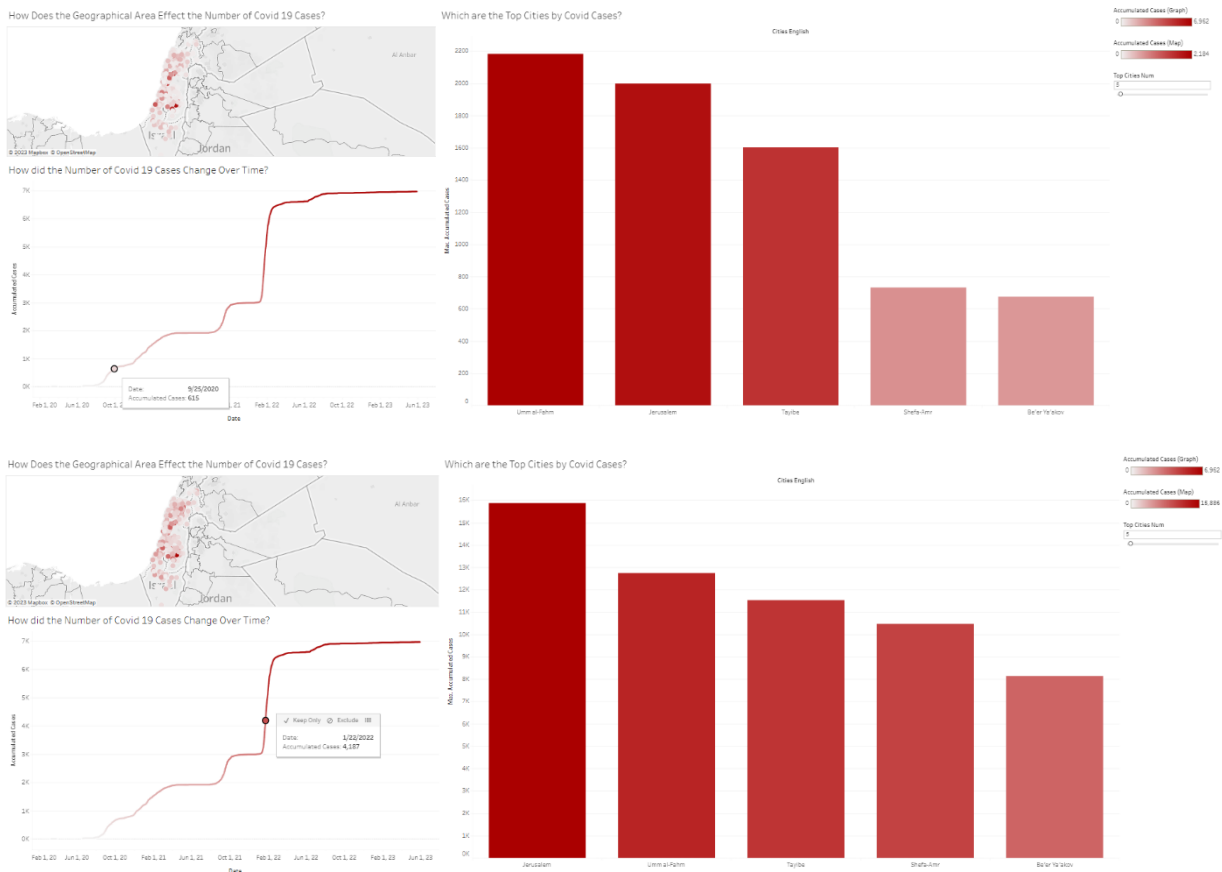
Top Cities Num
17
0

Accumulated Cases
0 25,700

- האינטראקציה העיקרית בוויזואליזציה היא בין הערים לבין כמות החולים (ובין כל עיר לעיר)
 - יש קשר אינטראקטיבי בין וויזואליזציה זו לבין האחרות – ניתן לבחור רגע זמן מסוים בוויזואליזציית הגרף Cases vs Time, והגרף משתנה כך שמציג את האזורים עם הכי הרבה מקרי קורונה לפי נקודת הזמן שנבחרה.
 - בנוסף, ניתן לבחור את האזורים שרוצים להציג בצורה אינטראקטיבית.
- Task של וויזואליזציה זו: Discover Outliers

- action של וויזואליזציה זו היא Discover – הוויזואליזציה נותנת למשתמש לבחור רגעי זמן ולבחור את הא שדרכו הוא רוצה לראות את המידע. אין בהכרח התמקדות בהשוואה בין רגעים או בין א-ים שונים.
- target של וויזואליזציה זו היא Outliers – המטרה העיקרית של הוויזואליזציה היא הצגת הערים עם הכי הרבה מקרי מחלה, במילים אחרות נקודות הקצה, שמהוות את מוקדי המחלה במדינה.

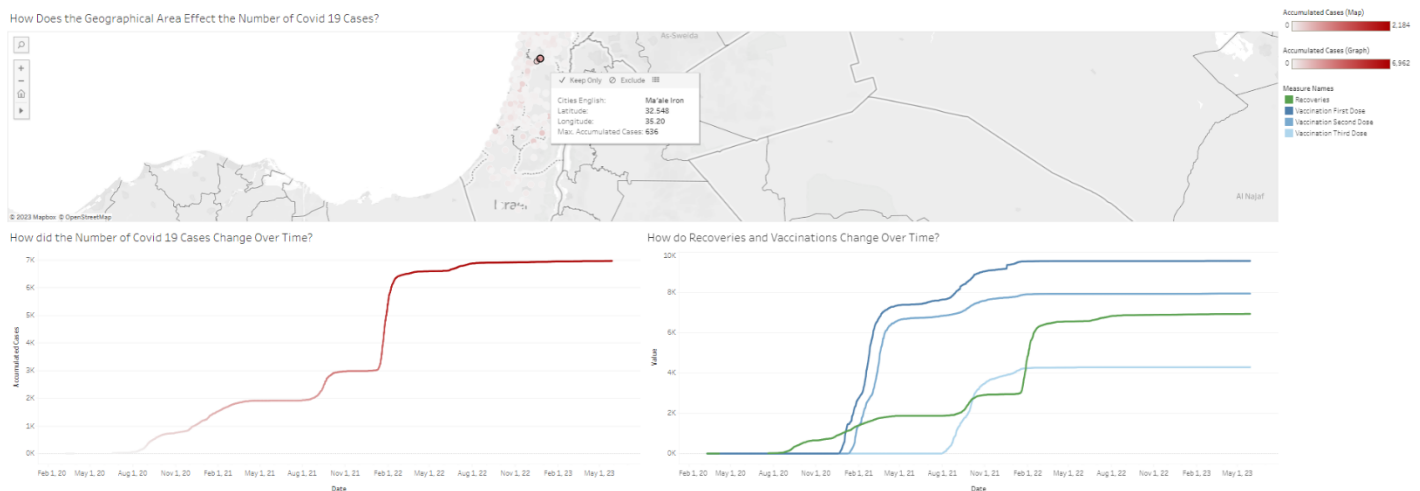
- dashboard זה מחולק לשלושה חלקים והוא מראה מה הקשר בין מספר החולים לפי האספקט של אזור גיאוגרפי לבין האספקט של הזמן. כלומר, ניתן לסנן את המידע בו לפי נקודת זמן. ניתן לבחור תאריך מסוים ולראות את השינויים לפי תאריך זה. למשל, ניתן לראות משתי הדוגמאות מטה שבתחילת המגיפה מספר מקרי הקורונה באום אל פחם היה יותר גדול ממספר מקרי הקורונה בירושלים אך לאורך הזמן מגמה זו השתנתה.



- האינטראקציה העיקרית בוויזואליזציה היא בין כמות החולים, זמן ומיקום (עיר). ההתמקדות העיקרית של הוויזואליזציה היא בזמן.

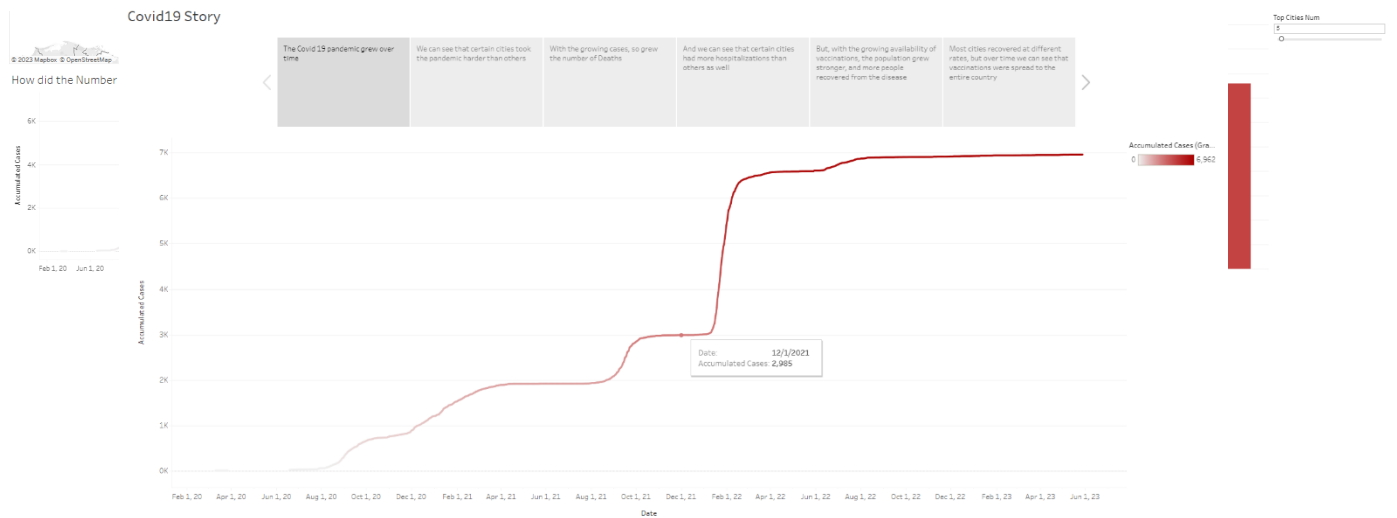
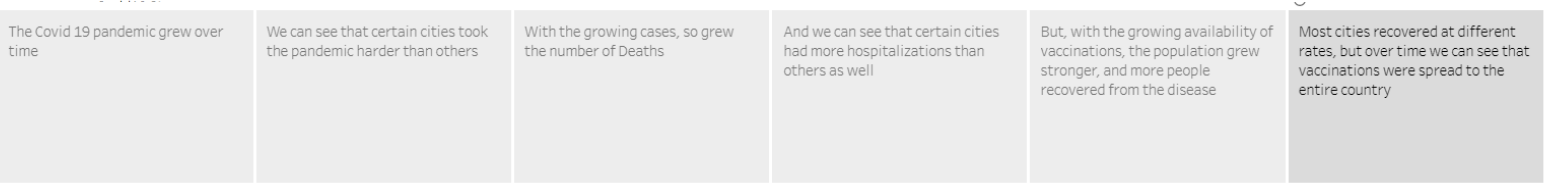
- ניתן לבחור רגע זמן מסוים בוויזואליזציית הגרף Cases vs Time, וויזואליזציות המפה והbar chart משתנים בהתאם לזמן שנבחר.
- בנוסף, ניתן לבחור את האזור עבור הbar chart.
- Task של וויזואליזציה זו: Present Trends
 - ההחלטות של וויזואליזציה זו היא Present – הוויזואליזציה מציגה את הנתונים למשתמש ללא צורך מיוחד באינטראקטיביות (למרות שזו תורמת לוויזואליזציה להציג את המידע). היא מציגה את המידע בכמה מובנים שונים ומספרת סיפור כללי.
 - הtarget של וויזואליזציה זו היא Trends – המטרה העיקרית של הוויזואליזציה היא השינוי של הנתונים לאורך זמן, במילים אחרות הtrend של כמות מקרי המחלה לאורך זמן ולפי מיקום.

- ב-dashboards זה, נוכל לבחור כל עיר ולבחון כל עיר בפני עצמה. גרף זה נותן למשתמש את האופציה לבחור עיר ספציפית והתסכל רק עליה. גרף זה הוא גם אחד הגרפים אשר פתר לנו את בעיית הנרמול מכיוון שבגרף זה מסתכלים על כל עיר באופן פרטני. גרף ממוקדם יותר. כלומר, בגרף זה ניתן לסנן את המידע על ידי בחירת עיר במפה ולראות את השינוי במספר חולי הקורונה, המחלימים והמתחשנים לאורך זמן.

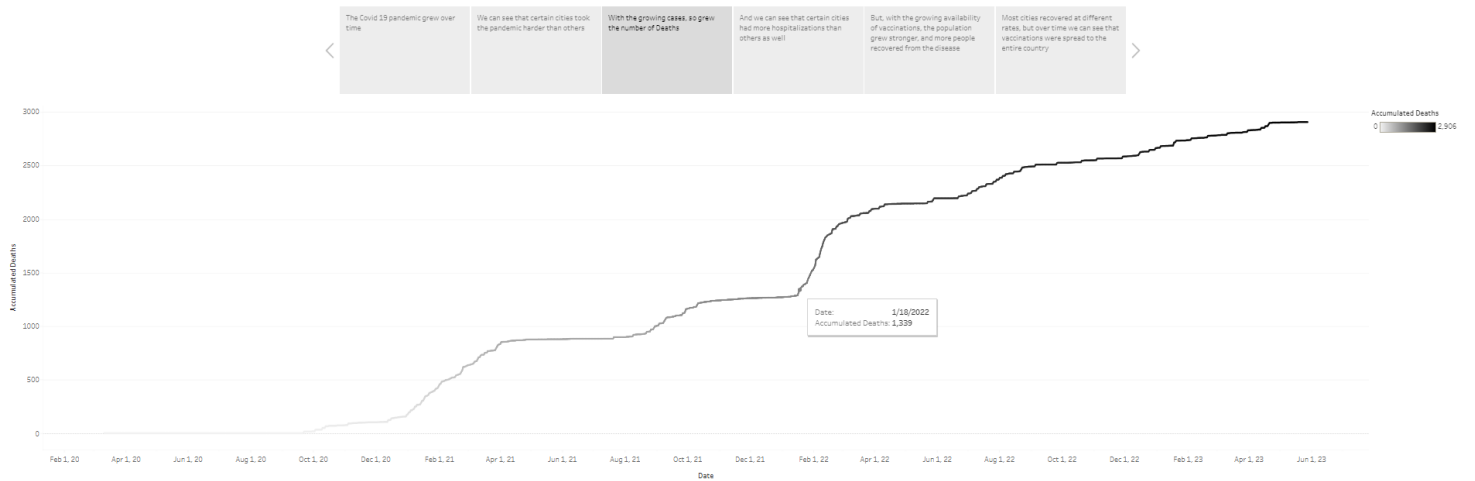


- המיפויים הם כמו שתיארנו בגרפים הבודדים.
- האינטראקציה העיקרית בוויזואליזציה היא בין מיקום וזמן לבין כמות מקרי המחלה, החיסונים וההבראה. ההתמקדות העיקרית של הוויזואליזציה היא במיקום.
 - ניתן לבחור מיקום מסוים על המפה, ובכך לסנן את הגרפים שייציגו את השינוי לאורך זמן עבור העיר שנבחרה.
- Task של וויזואליזציה זו: Lookup Trends
 - הaction של וויזואליזציה זו היא Lookup – הוויזואליזציה מציגה את הנתונים למשתמש לפי מיקום וזמן ההתמקדות העיקרית שלה. הבחירה היא בעיר מסוימת מאזור מסוים ששניהם ידועים מראש למטרת הבחירה, ולכן מדובר בlookup.
 - הtarget של וויזואליזציה זו היא Trends – המטרה העיקרית של הוויזואליזציה היא השינוי של הנתונים לאורך זמן, במילים אחרות הtrend של מקרי המחלה, החיסונים וההבראה לאורך זמן, בהתאם לבחירת הlookup.

• Story זה מאגד את כל הפרויקט עם הסברים נוספים הממחישים רצף של זמן ושינויים לאורך זמן.

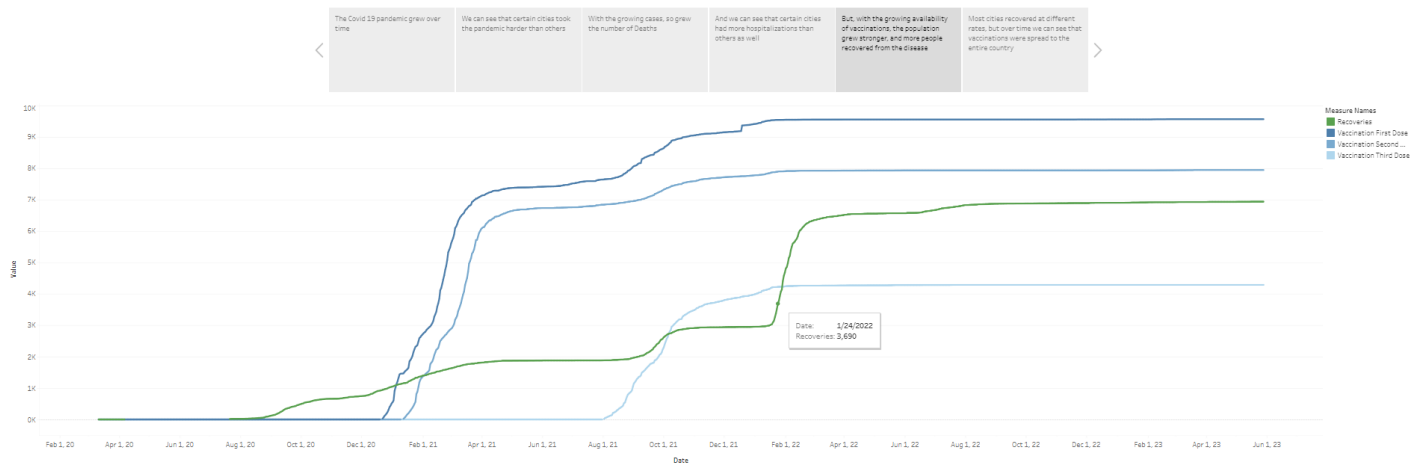


Covid19 Story





Covid19 Story



Covid19 Story

