

פרויקט קורס – Data Analytics

בפרויקט זה תבצעו ניתוח אנליטי ותבנו Story עבור נתוני הזמנות של שיט תעוגות. מרבית ההנחיות מבוססות על תוכנות של מדרתם במהלך הסמסטר. עם זאת הושתלו בפרויקט גם מספר דרישות שאתם נדרשים לחפש פתרון/לימוד בלבד על מנת לממשן.

nymo לב – את הפרויקט יש לבצע בקבוצות של בין 3 ל 4 סטודנטים. במידה ולא תמצאו קבוצה עד לסוף הסמסטר, פנו אליו ואני אוצר קבוצות על בסיס סטודנטים שלא מצאו שותפים. יודגש כי **לא תאושר** הגשה של קבוצה בת פחות מ 3 סטודנטים.

בספריית הפרויקט סופק לכם קובץ בשם reservations.csv.

בקובץ כל שורה מייצגת הזמנה של לקוחות לשיט בספינות תעוגות. להלן הסבר על מספר שדות בקובץ:

-	StatD: התאריך הרלוונטי לעדכון הסטטוס האחרון לשיט
-	ExtraLock: מספר הлокרים הנוספים שהזמננו על ידי הלוקן לאחסון חפצים
-	Special: מספר הבקשות המיווחדות של הלוקן (למשל קומה בספינה, גישה ללאונג' ועוד)
-	stat: מצב הזמנה, למשל הזמנה מבוטלת, הלוקן כבר עשה צ'קআট, הלוקן לא הופיע בזמן
-	CT: סוג הלוקן בין אם מדובר בлокן מזדמן, קבוצה, חוות, ועוד
-	AgentNr: מזהה הסוכן שביצע את הזמנה עבור הלוקן
-	Cmpy: מזהה החברה שבהעסקה הייתה בראשית המתנה טרם האישור שלא עבור הלוקן
-	waiting: מספר הימים שהעסקה הייתה בראשית המתנה טרם האישור שלא עבור הלוקן
-	DailyAVG: התעריף הממוצע ליום עבור העסקה (долרים)
-	ExtraLock: מספר הлокרים הנוספים שהזמננו על ידי הלוקן לאחסון חפצים
-	Special: מספר הבקשות המיווחדות של הלוקן (למשל קומה בספינה, גישה ללאונג' ועוד)
-	stat: מצב הזמנה, למשל הזמנה מבוטלת, הלוקן כבר עשה צ'קআট, הלוקן לא הופיע בזמן
-	Country: מדינת המקור של מזמין השיט
-	market: סגמנט השוק בין אם מדובר בהזמנה דרך אינטרנט (Online TA), חברתית (Corporate), ישירה (Direct), סוכנות תיירות ללא אתר ועוד
-	Food: סוג חבילה הארוחה במידעה וקיימות בהזמנה. BB – Bed & Breakfast .HB- Half Board .NBA.NB – NBA.FB – Full board
-	Weekendnights, weeknights: מספר הלילות בזמן השיט שמהווים סופ"ש ומספר הלילות שמהווים ימים באמצעות השבוע.
-	DOM: מספר הימים בחודש השיט
-	M: חדש השיט
-	Y: שנת השיט
-	TTT: מספר הימים בין תאריך הזמנה לתאריך השיט
-	cancel: האם הזמנה בוטלה
-	(Deluxe/Regular) Cruise: סוג השיט
-	returning: האם מדובר בлокן חוזר
-	prevNotCanc,prevCanc : מספר הזמנות שהלוקן זיהה ביטול או לא ביטל בעבר
-	reservedRT: קוד סוג החדר שהטפסו של דבר הוקצה ללקוח (לפעמים יהיה שונה מזוינה reserved כי חסרים חדרים מסווג מבודק, או שהלוקן ביקש לשנות כשללה לספינה)
-	assignedRT: קוד סוג החדר שהטפסו של דבר הוקצה ללקוח (לפעמים יהיה שונה מזוינה assigned כי מזהה לסוג ההזמנה שפועלו בהזמנה בין רגע ההזמנה הראשונית ועד יום השיט או הביטול deposit: מזהה לסוג ההזמנה מבחינת פקdon להבטחת החדר, והאם העסקה ניתנת לביטול ללא קנס או אם קנס).
-	AgentNr: מזהה הסוכן שביצע את הזמנה עבור הלוקן
-	Cmpy: מזהה החברה שבהעסקה הייתה בראשית המתנה טרם האישור שלא עבור הלוקן
-	waiting: מספר הימים שהעסקה הייתה בראשית המתנה טרם האישור שלא עבור הלוקן
-	CT: סוג הלוקן בין אם מדובר בлокן מזדמן, קבוצה, חוות, ועוד
-	DailyAVG: התעריף הממוצע ליום עבור העסקה (долרים)
-	ExtraLock: מספר הлокרים הנוספים שהזמננו על ידי הלוקן לאחסון חפצים
-	Special: מספר הבקשות המיווחדות של הלוקן (למשל קומה בספינה, גישה ללאונג' ועוד)
-	stat: מצב הזמנה, למשל הזמנה מבוטלת, הלוקן כבר עשה צ'קআট, הלוקן לא הופיע בזמן
-	StatD: התאריך הרלוונטי לעדכון הסטטוס האחרון

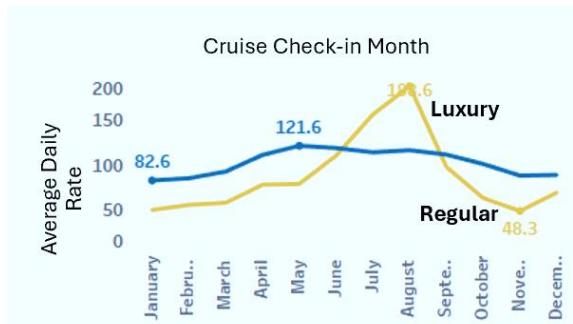
במשימה זו עליכם לבנות Story עבור נתוני הזמןות השיט. שימו לב – את ה Story אתם צריכים לתכנן בעצמכם מבחינת מראה, החלטה היקן להציג אלמנטים מסוימים ובנויות התוצאות כMOVED. עם זאת – עליכם לעמוד גם ברשימת הדרישות הטכניות הבאות:

- ה Story יוכל לפחות 5 דשبورדים. כל דשبورד יוכל לפחות 5 אובייקטים/תמונות גרפיות.
- מבין 5 הדשبورדים, 3 מהם יתיחסו לפחותם הבאים (טור מימוש מלא של הדרישות הבאות):

○ דשبورד הזמןות

- יציג מספר הזמןות המאושרות על כל הדטה (זמןנות שהשדה stat שלהן הוא (checkout)
- יציג סכום הזמןות במילוני דולרים על כל הדטה (שימו לב לסכום רק עבר'
- הזמןות שהסטטוס שלהם על צ'ק אוט או deposit שלו הוא non refundable (non refundable deposit)
- יציג סכום הזמןות המופסדות (זמןנות שהסטטוס שלהם שונה מ'צ'ק אוט' וגם שהוא שונה מ'non refundable deposit')
- יציג מספר האורחים שהו בשיט לאורך כל התקופה (שימו לב שיש להתייחס רק להזמןנות שהסטטוס שלהם הוא צ'ק אוט ולקחת בחשבון את המבוגרים, הילדים והתינוקות).
- יציג אחוז התפוצה הממוצע (يحושב על ידי חישוב אחוז הזמןות שהסטטוס שלהם הוא צ'ק אוט מתוך סך הזמןות) על פני כל התקופה
- יציג גרפ' שוק המציג את התפלגות ארצות/מדינות המקור של מזמינים השיט על פני כל הדטה
- יציג גרפ' אשר יציג את ממוצע התעריף היומי עבור כל חודש של שיט מבין 12 חודשים השנה (על פני כל התקופה) ועבור כל סוג שיט. פירוט: ציר X הוא 12 החודשים מינואר עד דצמבר, ציר Y הוא ממוצע התעריף היומי עבור הזמןות שהשיט שלהם מתחליל בחודש שמצוין בציר X, ובגרף יופיעו 2 עקומות – אחת עבור הזמןות של שיט מסוג Deluxe והשנייה עבור הזמןות של שיט מסוג Regular

• דוגמא:



- יציג אחוז התפוצה עבור חודש ספציפי (ניתן לבחירה על ידי המשתמש) וסוג השיט (ניתן לבחירה על ידי המשתמש גם כן) כר שיוופיע calendar של החודש המבוקש והימים בחודש יצבעו על בהתאם לאחוז התפוצה מבהיר(תפוצה נוכחית) לכמה (תפוצה גבוהה).
- דוגמא :

<input style="border: 1px solid #ccc; border-radius: 5px; padding: 2px 10px; width: 150px; height: 20px;" type="button" value="September 2016"/>	<input style="border: 1px solid #ccc; border-radius: 5px; padding: 2px 10px; width: 150px; height: 20px;" type="button" value="Luxury"/>							
Reservation Status Date								
Week of Reser.. Sunday Monday Tuesday Wedne.. Thursd.. Friday Saturd..								
Week 36								
Week 37	<table border="1" style="border-collapse: collapse; width: 100%;"> <tr><td style="width: 25px; text-align: center;">4</td><td style="width: 25px; text-align: center;">5</td><td style="width: 25px; text-align: center;">6</td><td style="width: 25px; text-align: center;">7</td><td style="width: 25px; text-align: center;">8</td><td style="width: 25px; text-align: center;">9</td><td style="width: 25px; text-align: center;">10</td></tr> </table>	4	5	6	7	8	9	10
4	5	6	7	8	9	10		
Week 38	<table border="1" style="border-collapse: collapse; width: 100%;"> <tr><td style="width: 25px; text-align: center;">11</td><td style="width: 25px; text-align: center;">12</td><td style="width: 25px; text-align: center;">13</td><td style="width: 25px; text-align: center;">14</td><td style="width: 25px; text-align: center;">15</td><td style="width: 25px; text-align: center;">16</td><td style="width: 25px; text-align: center;">17</td></tr> </table>	11	12	13	14	15	16	17
11	12	13	14	15	16	17		
Week 39	<table border="1" style="border-collapse: collapse; width: 100%;"> <tr><td style="width: 25px; text-align: center;">18</td><td style="width: 25px; text-align: center;">19</td><td style="width: 25px; text-align: center;">20</td><td style="width: 25px; background-color: #90EE90; text-align: center;">21</td><td style="width: 25px; text-align: center;">22</td><td style="width: 25px; background-color: #90EE90; text-align: center;">23</td><td style="width: 25px; text-align: center;">24</td></tr> </table>	18	19	20	21	22	23	24
18	19	20	21	22	23	24		
Week 40	<table border="1" style="border-collapse: collapse; width: 100%;"> <tr><td style="width: 25px; text-align: center;">25</td><td style="width: 25px; text-align: center;">26</td><td style="width: 25px; text-align: center;">27</td><td style="width: 25px; text-align: center;">28</td><td style="width: 25px; text-align: center;">29</td><td style="width: 25px; text-align: center;">30</td></tr> </table>	25	26	27	28	29	30	
25	26	27	28	29	30			

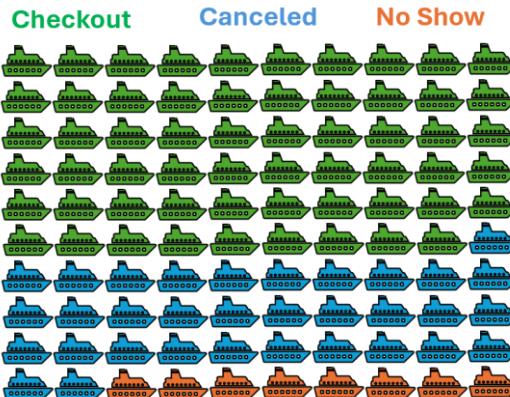
ויצג גרף bar המציג את מספר ההזמנות לפי חלוקה ל:

- מכילות רק ימים באמצע השבוע

- מכילות רק ימים בסוף השבוע

- מכילות ימים גם בסוף השבוע וגם באמצע השבוע

ויצג יציג ויזואלי של אחוז ההזמנות על פי הסטטוסים השונים (הזמןות רגילות, ביטולים, אי הופעה) כהמחשה מתוך 100 "ספינות". למשל אם 69 אחוז מהזמןות עברו צ'ק אאוט, 33 אחוז עברו ביטול ו 8 אחוז לא הופיעו לשיט צה יראה כך:



- לדוגמה:

o דשبورד הכנסות

▪ יהיה אפשר לעبور בין 2 ייצוגים ויזואליים (באמצעות כפתור או tab) עברו מדיניות מקור על פי הכנסות:

▪ ויצג יציג ויזואלי של טבלה המציג את 10 מדיניות המקור שעבור הכנסה המוצברת לאורך כל התקופה הניבוי את סכום הכנסה המוצברת הגובה ביותר. יש להציג את שמות 10 המדיניות ואת הסכום המוצבר עבור כל אחת.

▪ בנוסף לטבלה, יציג גם ייצוג נוספים של מפת העולם, שהמדיניות בה צבועות לפי סכום הכנסה המוצברת עבור אותה מדינת מקור.

▪ ויצג גרף סirk המראה את החלוקה לשגננטים שונים של השוק על פי הכנסה המוצברת לאורך כל התקופה

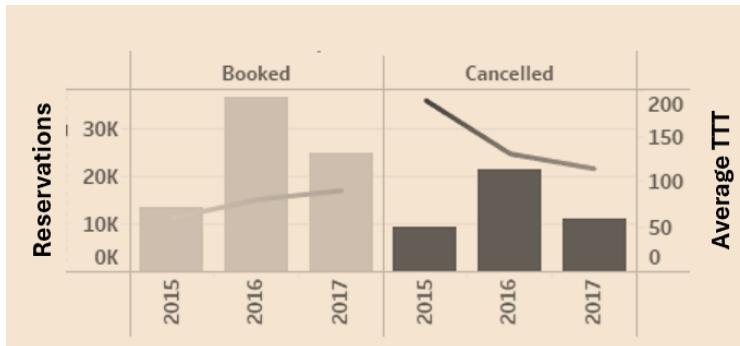
▪ ויצג גרף עבור 12 חודשים השנה מבחינת הכנסות והפסדים (מצטבר לאותו החודש שבו היה צ'ק אין לשיט בכל השנים של הדטה).

▪ הכנסות יחושבו ויוצגו במיליאני דולרים על כל הדטה (שים לב לסכום רק עבור הזמןות שהסתטוו שלhn על צ'ק אאוט או deposit שלhn הוא חסן refundable)

▪ הפסדים יחושבו ויוצגו באלפי דולרים עבור הזמןות המופסדות (הזמןות

שהסתטוו שלhn שונה מצ'ק אאוט **וגם** שהdeposit שלhn שונה מ חסן refundable)

- את הכנסות והפסדים יש להציג אחד מתחת לשני (הגרפים לא עולים אחד על השני – כלומר אין כאן secondary axis אלא הגרפים אחת מתחת לשני)
 - יציג גרף סוֹק עבורי הציגונות המוצטברות עבור סוג השיט (Deluxe/Regular). יש לציין בגרף גם את האחוזים של כל סוג שיט על פי סכום הכנסות המוצטבר וכך וכך את הסכומים עצם ב מיליון דולר.
 - יציג גרף עבור 12 חודשים השנה מבחינת מספר לקוחות והכנסות המוצטברות (מצטבר לאותו החודש שבו היה צ'ק אין לשיט בכל השנים של הדטה). את הגרפים הפנימיים עבור מספר לקוחות והכנסות המוצטברות יש להציג אחד על השני (גרף עם secondary axis).
- **דשبورד ביטולי ל��וחות**
 - יציג מקטע המכיל 3 גרפים של **סיוֹק המראים את אחוז הזמין שבוטל ושללא בוטל** עבור 3 סוגי לקוחות: **יחידים**(הזמןה של לקוח מבוגר בודד ללא ילדים ותינוקות), **זוגות** (הזמןה של 2 לקוחות מבוגרים ללא ילדים ותינוקות) ו**קובזות** (כל קומבינציה אחרת).
 - יציג גרף המציג את התפלגות הזמין שבוטל לפי ימי השבוע עבור הזמין **مبוטלות** ועבור הזמן **שאין מבוטלות**. ממוצע זמן ה-ZTT ומספר הזמין עבור ההזמין מבוטלות יוצגו כ bar. ה אחד הוא bar ישבו אחד על השני (שים לב שישנו ויזואליות 2 מקטעים לגרף – הימני עבור ההזמין המבוטלות והשמאלי עבור הזמן שבוטל).
 - דוגמא:



- יציג גרף המציג את אחוז ההזמין שבוטל ושללא בוטל עבור:
 - לקוחות שלא היו ברשימה המתנה
 - היו ברשימה המתנה עד שבוע
 - היו ברשימה המתנה בין שבוע לחודש
 - היו ברשימה המתנה בין חודש ל 3 חודשים
 - היו ברשימה המתנה יותר מ 3 חודשים
- יציג גרף המציג את אחוז הביטולים ואחוז התפוצה (אחד מעל השני) עבור כל אחד מהסוגניטים השונים הקיימים בשוק

○ **שילוב תחזית באחד הדשبورדים האחרים**

- 3 הדשبورדים המפורטים לעיל (**הזמין, הכנסות, ביטולי ל��וחות**) אמורים להכיל את התכנים שצינו בכל אחד מהם. מלבד הדשبورדים האלה, עליהם למשם לפחות 2 דשبورדים נוספים (כך שייהי סה"כ 5 דשبورדים). באחד מהdashboards שתיצרו, עליהם לשלב גרף המציג תחזית עתידית (פרדייקציה) עבור הזמן שבוטלים. ציר X יציג את החודשים של הזמן שבוטלים ובציר Y יופיעו גרפים (אוחנו) עבור הזמן שבוטלים ברמה החודשית. עליהם לספק תחזית של שנה קדימה מעלה הדטה הקיימת. לדוגמא בגרף הבא (לא קשור לדטה שלנו) בכחול מופיעים ערכים שנמצאים בקבוץ ובכתום מופיעים ערכי התחזית של השנה הבאה.



תוכלו לקרוא כיצד לספק תחזיות באמצעות טאבלו כאן:

https://help.tableau.com/current/pro/desktop/en-us/forecast_create.htm

- מעבר לדרישות - תרגישו חופשיים לחושב מחוץ לקופסה לגבי הנתונים שקיבלתם, ולנסות ליצור דשborדים יצירתיים – ככלו המבאים ערך עבור מקבל החלטות, עם תובנות עסקיות על הדטה.

- ניתוח מבוסס פיתון

- עיקר העבודה שקיבלתם מבוססת על טאבלו, אך במהלך הקורס קיבתם גם סקירה ראשונית על פיתון ועובדת עם Pandas והוצאות גרפים בסיסיים. בסעיף זה עליכם לבחור 4 גرافים מתוך הדרישות שמופיעות לעיל (לבחירתכם) ולמש אתם מחוץ לטאבלו באמצעות Jupyter Notebook.

- הגשה

- את הפרויקט יש להגיש עד לתאריך שיפורסם על גבי המטלה במודול.
- במסגרת ההגשה תתבקשו מלא טופס המכיל תוצאות מתוך העבודה שביצעתם, יש להקפיד למלא את כל הפרטים ולהעלות screenshots כפי שתתבקשו.
- לצורך הגשת קוד המקור (גם של טאבלו וגם של פיתון) תפתח תיבת הגשה. שמות הקבצים יהיו לפי שמות הסטודנטים ותעדות הזהות של כל חבר הקבוצה.

ב ה צ ל ח ה !