

פרויקט קורס – Data Analytics

בפרויקט זה תבצעו ניתוח אנליטי ותבנו Story עבור נתוני הזמנות של שייט תענוגות. מרבית ההנחיות מבוססות על תכונות שלמדתם במהלך הסמסטר. עם זאת הושתלו בפרויקט גם מספר דרישות שאתם נדרשים לחפש פתרון/ללמוד לבד על מנת לממשן.

שימו לב – את הפרויקט יש לבצע בקבוצות של **3 סטודנטים**. במידה ולא תמצאו קבוצה עד לסוף הסמסטר, פנוי אלי ואני אצור קבוצות על בסיס סטודנטים שלא מצאו שותפים.

בספריית הפרויקט סופק לכם קובץ בשם reservations.csv.

בקובץ כל שורה מייצגת הזמנה של לקוחות לשייט בספינת תענוגות. להלן הסבר על מספר שדות בקובץ:

- Cruise: סוג השייט (Deluxe/Regular)
- cancel: האם ההזמנה בוטלה
- TTT: מספר הימים בין תאריך ההזמנה לתאריך השייט
- Y: שנת השייט
- M: חודש השייט
- W: מספר השבוע בשנה של השייט
- DOM: מספר היום בחודש השייט
- Weekendnights, weeknights: מספר הלילות בזמן השייט שמהווים סופ"ש ומספר הלילות שמהווים ימים באמצע השבוע.
- ADLT, CHLD, BBY: מספר המבוגרים, ילדים ותינוקות בחדר שהוזמן בספינה
- Food: סוג חבילת הארוחה במידה וקיימת בהזמנה. BB – Bed & Breakfast .HB- Half Board .NBA.NB .FB – Full board אין ארוחה מוזמנת מראש (מזמינים באופן מזדמן במקום).
- Country: מדינת המקור של מזמין השייט
- market: סגמנט השוק בין אם מדובר בהזמנה דרך אתר (Online TA), חברה (Corporate), ישירה (Direct), סוכנות תיירות ללא אתר ועוד
- channel: ערוץ ההפצה שממנו מגיעה העסקה. TA משמעותו Travel Agents, TO משמעותו Tour Operators וכד'
- returning: האם מדובר בלקוח חוזר
- prevNotCanc, prevCanc: מספר ההזמנות שהלקוח של הזמנה זו ביטל או לא ביטל בעבר
- reservedRT: קוד סוג החדר שלקוח הזמין (מופיע כקוד מסיבות אנונימיות)
- assignedRT: קוד סוג החדר שבסופו של דבר הוקצה ללקוח (לפעמים יהיה שונה מreserved כי חסרים חדרים מסוג מבוקש, או שהלקוח ביקש לשנות כשעלה לספינה)
- changes: מספר השינויים שבוצעו בהזמנה בין רגע ההזמנה הראשון ועד יום השייט או הביטול
- deposit: מזהה לסוג ההזמנה מבחינת פקדון להבטחת החדר, והאם העסקה ניתנת לביטול ללא קנס או אם קנס.
- AgentNr: מזהה הסוכן שביצע את ההזמנה עבור הלקוח
- Cmpy: מזהה החברה שביצעה את ההזמנה במידה ומדובר בחברה
- waiting: מספר הימים שהעסקה היתה ברשימת המתנה טרם האישור שלה עבור הלקוח
- CT: סוג הלקוח בין אם מדובר בלקוח מזדמן, קבוצה, חוזה, וכד'
- DailyAVG: התעריף הממוצע ליום עבור העסקה (דולרים)
- ExtraLock: מספר הלוקרים הנוספים שהוזמנו על ידי הלקוח לאחסון חפצים
- Special: מספר הבקשות המיוחדות של הלקוח (למשל קומה בספינה, גישה ללאונג' וכד')
- stat: מצב ההזמנה, למשל הזמנה מבוטלת, הלקוח כבר עשה צ'ק אאוט, הלקוח לא הופיע בזמן לשייט
- StatD: התאריך הרלוונטי לעדכון הסטטוס האחרון

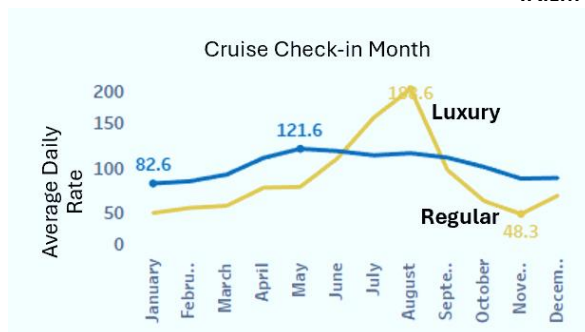
במשימה זו עליכם לבנות Story עבור נתוני הזמנות השייט. שימו לב – את ה Story אתם צריכים לתכנן בעצמכם מבחינת מראה, החלטה היכן להציג אלמנטים מסוימים ובניית התצוגות כמובן. עם זאת – עליכם לעמוד גם ברשימת הדרישות הטכניות הבאות:

- ה Story יכיל לפחות 5 דשבורדים. כל דשבורד יכיל לפחות 5 אובייקטים/תצוגות גרפיות.
- מבין 5 הדשבורדים, 3 מהם יתייחסו כל אחד לנושאים הבאים (תוך מימוש מלא של הדרישות הבאות):

○ הזמנות

- יוצג מספר ההזמנות המאושרות על כל הדאטה (הזמנות שהשדה stat שלהן הוא checkout)
- יוצג סכום ההזמנות במיליוני דולרים על כל הדאטה (שימו לב לסכום רק עבור הזמנות שהסטטוס שלהן על צ'קאאוט **או** ה deposit שלהן הוא non refundable)
- יוצג סכום ההזמנות המופסדות (הזמנות שהסטטוס שלהן שונה מצ'ק אאוט **וגם** שה deposit שלהן שונה מ non refundable)
- יוצג מספר האורחים ששהו בשייט לאורך כל התקופה (שימו לב שיש להתייחס רק להזמנות שהסטטוס שלהן הוא צ'ק אאוט ולקחת בחשבון את המבוגרים, הילדים והתינוקות).
- יוצג אחוז התפוסה הממוצע (יחושב על ידי חישוב אחוז ההזמנות שהסטטוס שלהן הוא צ'ק אאוט מתוך סך ההזמנות) על פני כל התקופה
- יוצג גרף pie המציג את התפלגות ארצות/מדינות המקור של מזמיני השייט על פני כל הדאטה
- יוצג גרף אשר יציג את ממוצע התעריף היומי עבור כל חודש של שייט מבין 12 חודשי השנה (על פני כל התקופה) ועבור כל סוג שייט. פירוט: ציר X הוא 12 החודשים מינואר עד דצמבר, ציר Y הוא ממוצע התעריף היומי עבור הזמנות שהשייט שלהן מתחיל בחודש שמצוין בציר X, ובגרף יופיעו 2 עקומות – אחת עבור הזמנות של שייט מסוג Deluxe והשניה עבור הזמנות של שייט מסוג Regular

• דוגמא:



- יוצג אחוז התפוסה עבור חודש ספציפי (ניתן לבחירה על ידי המשתמש) וסוג השייט (ניתן לבחירה על ידי המשתמש גם כן) כך שיופיע calendar של החודש המבוקש והימים בחודש ייצבעו על בהתאם לאחוז התפוסה מבהיר(תפוסה נמוכה) לכה (תפוסה גבוהה).

• דוגמא :

September 2016

Luxury

Reservation Status Date

Week of Reser..

Sunday

Monday

Tuesday

Wedne..

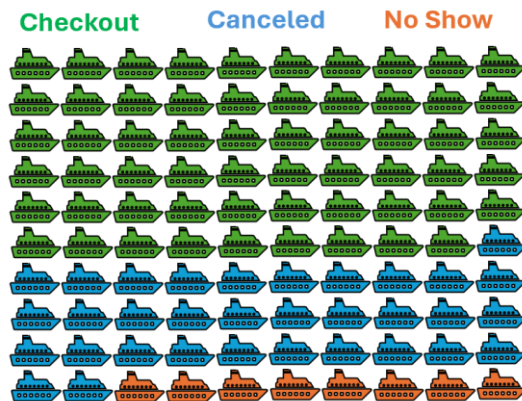
Thursd..

Friday

Saturd..

Week 36					1	2	3
Week 37	4	5	6	7	8	9	10
Week 38	11	12	13	14	15	16	17
Week 39	18	19	20	21	22	23	24
Week 40	25	26	27	28	29	30	

- יוצג גרף bar המציג את מספר ההזמנות לפי חלוקה ל:
 - מכילות רק ימים באמצע השבוע
 - מכילות רק ימים בסופי שבוע
 - מכילות ימים גם בסופי שבוע וגם מאמצע השבוע
- יוצג ייצוג ויזואלי של אחוז ההזמנות על פי הסטטוסים השונים (הזמנות רגילות, ביטולים, אי הופעה) כהמחשה מתוך 100 "ספיות". למשל אם 69 אחוז מההזמנות עברו צ'ק אאוט, 33 אחוז עברו ביטול ו 8 אחוז לא הופיעו לשייט זה יראה כך:



• לדוגמא:

○ הכנסות

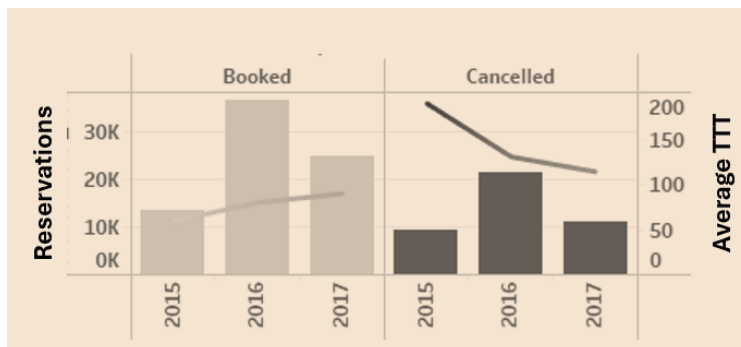
- יהיה אפשר לעבור בין 2 ייצוגים ויזואליים (באמצעות כפתור או tab) עבור מדינות מקור על פי הכנסות:
 - יוצג ייצוג ויזואלי של טבלה המציג את 10 מדינות המקור שעבור ההכנסה המצטברת לאורך כל התקופה הניבו את סכום ההכנסה המצטברת הגבוה ביותר. יש להציג את שמות 10 המדינות ואת הסכום המצטבר עבור כל אחת.
 - בנוסף לטבלה, יוצג גם ייצוג נוסף של מפת העולם, שהמדינות בה צבעות לפי סכום ההכנסה המצטברת עבור אותה מדינת מקור.
- יוצג גרף pie המראה את החלוקה לסגמנטים שונים של השוק על פי ההכנסה המצטברת לאורך כל התקופה
- יוצג גרף עבור 12 חודשי השנה מבחינת הכנסות והפסדים (מצטבר לאותו החודש שבו היה צ'ק אין לשייט בכל השנים של הדאטה).
- **הכנסות** יחושבו ויוצגו **במיליוני דולרים** על כל הדאטה (שימו לב לסכום רק עבור הזמנות שהסטטוס שלהן על צ'קאאוט או deposit שלהן הוא non refundable)
- **הפסדים** יחושבו ויוצגו **באלפי דולרים** עבור ההזמנות המופסדות (ההזמנות שהסטטוס שלהן שונה מצ'ק אאוט וגם deposit שלהן שונה מ non refundable)

- את הכנסות וההפסדים יש להציג אחד מתחת לשני (הגרפים לא עולים אחד על השני – כלומר אין כאן secondary axis אלא הגרפים אחת מתחת לשני)
- יוצג גרף pie עבור ההזמנות המצטברות עבור סוג השייט (Deluxe/Regular). יש לציין בגרף גם את האחוזים של כל סוג שייט על פי סכום ההכנסות המצטבר וכמו כן את הסכומים עצמם במיליוני דולרים.
- יוצג גרף עבור 12 חודשי השנה מבחינת מספר לקוחות והכנסות המצטברות (מצטבר לאותו החודש שבו היה צ'ק אין לשייט בכל השנים של הדאטה). את הגרפים הפנימיים עבור מספר הלקוחות וההכנסות המצטברות יש להציג אחד על השני (גרף עם secondary axis)

○ ביטולי לקוחות

- יוצג מקטע המכיל 3 גרפים של pie המראים את אחוז ההזמנות שבוטלו ושלא בוטלו עבור 3 סוגים של לקוחות: **יחידים** (הזמנה של לקוח מבוגר בודד ללא ילדים ותינוקות), **זוגות** (הזמנה של 2 לקוחות מבוגרים ללא ילדים ותינוקות) ו**קבוצות** (כל קומבינציה אחרת).
- יוצג גרף המציג את התפלגות ההזמנות המבוטלות לפי ימי השבוע
- יוצג גרף ברזולוציה **שנתית** המציג את **ממוצע זמני ה TTT** ואת **מספר הזמנות** עבור הזמנות **מבוטלות** ועבור הזמנות **שאינן מבוטלות**. ממוצע זמני ה TTT יוצגו כ line ומספר ההזמנות יוצגו כ bar. ה line וה bar ישבו אחד על השני (שימו לב שישנם ויזואליות 2 מקטעים לגרף – הימני עבור ההזמנות המבוטלות והשמאלי עבור ההזמנות שאינן מבוטלות).

• דוגמא:



- יוצג גרף המציג את אחוז ההזמנות שבוטלו ושלא בוטלו עבור:
 - לקוחות שלא היו ברשימת המתנה
 - היו ברשימת המתנה עד שבוע
 - היו ברשימת המתנה בין שבוע לחודש
 - היו ברשימת המתנה בין חודש ל 3 חודשים
 - היו ברשימת המתנה יותר מ 3 חודשים
- יוצג גרף המציג את אחוז הביטולים ואחוז התפוסה (אחד מעל השני) עבור כל אחד מהסגמנטים השונים הקיימים בשוק

○ שילוב תחזית באחד הדשבורדים האחרים

- 3 הדשבורדים המפורטים לעיל (**הזמנות, הכנסות, ביטולי לקוחות**) אמורים להכיל את התכנים שצוינו בכל אחד מהם. מלבד הדשבורדים האלה, עליכם לממש לפחות 2 דשבורדים נוספים (כך שיהיו סה"כ 5 דשבורדים). באחד מהדשבורדים שתיצרו, עליכם לשלב גרף המציג תחזית עתידית (פרדיקציה) עבור הזמנות וביטולים. ציר X ייצג את החודשים של ההזמנות לאורך השנים ובציר Y יופיעו גרפים (line) עבור ההזמנות והביטולים ברמה החודשית. עליכם לספק תחזית של שנה קדימה מעל הדאטה הקיים. לדוגמא בגרף הבא (לא קשור לדאטה שלנו) בכחול מופיעים ערכים שנמצאים בקובץ 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 לחשוב מחוץ לקופסה לגבי הנתונים שקיבלתם, ולנסות ליצור דשבורדים יצירתיים – כאלו המביאים ערך עבור מקבל ההחלטות, עם תובנות עסקיות על הדאטה.

- ניתוח מבוסס פייתון

- עיקר העבודה שקיבלתם מבוססת על טאבלו, אך במהלך הקורס קיבלתם גם סקירה ראשונית על פייתון ועבודה עם Pandas והוצאת גרפים בסיסיים. בסעיף זה עליכם לבחור 4 גרפים מתוך הדרישות שמופיעות למעלה (לבחירתכם) ולממש אותם מחוץ לטאבלו באמצעות Jupyter Notebook.

- הגשה והגנה

- את הפרויקט יש להגיש עד הערב שלפני יום ההגנה שלכם. תאריכים ההגנה מתפרסמים יחד עם תאריכי הבחינות על ידי המזכירות.
- ההגשה תהיה בקובץ zip שיכיל 2 קבצים- קובץ של טאבלו וקובץ של ipynb (מחברת יופיטר לפייתון). שמות הקבצים יהיו לפי שמות הסטודנטים ותעודות הזהות של כל חברי הקבוצה.
- במהלך ההגנה עליכם להציג את מה שיצרתם ולענות על שאלות שתישאלו. השאלות מתייחסות בעיקר למה שביצעתם בפרויקט (בין אם התבקשתם ובין אם בחרתם/זמתם) אך לא רק – אלא גם לחומר שנלמד במהלך הסמסטר שעליכם לשלוט בו לקראת ההגנה. כל סטודנט החבר בקבוצת הפרויקט צריך להכיר היטב את מה שנעשה על ידי כל אחד מחברי הקבוצה.

בהצלחה!