**שאלות למר עזריאלי**

שלום רב מר עזריאלי, אשמח לשאול מספר שאלות:

שאלות בנוגע לתעריפי נסיעות:

1. האם הnight bonus עבור נסיעות מעל 200 קילומטר מחושב על התעריף לאחר הוספת הextra milage עבור כל קילומטר אחרי 200? או שמחשבים אותו על מספר הקילומטרים שעברת ואז מוסיפים את ההפרש בין extra milage לבין תעריף של קילומטר רגיל? לא
2. אותה שאלה עבור weekend bonus ......
3. האם מחשבים קודם weekend bonus ואחרי זה על התוצאה מוסיפים extra milage או שקודם מחשבים extra milage ואז את הweekend bonus? קודם מחשב milage ואז אקסטרה
4. האם מחשבים גם weekend bonus וגם night bonus אם נוסעים בסופש בלילה? מחברים את האחוזים ואז מפעילים על הסכום הנוכחי
5. מה זה Weekend? רק מכניסת שבת ועד יציאת שבת? 16:00 בשישי עד 20:00 בשבת
6. מתי זה לילה? 22:00 עד 6:00

שאלות בנוגע למידע עצמו:

1. בקובץ **taarif.csv** חסרים מספר פרטים וחלקם נראים לא הגיוניים

בעיות:

* 1. ישנם לקוחות (customers) אשר משלמים 0 שקלים בונוס על נסיעה לילית (night bonus) או על נסיעה בסוף השבוע (weekend bonus). האם הדבר תקין ואינו טעות? תיקן אותי ועשינו 20 אחוז
  2. בשורה שבה הלקוח(customer) הינו electricity נראה שהבונוס לסוף שבוע לא הגיוני – 70,000 אחוז בונוס על כל קילומטר. חסר לו מידע על החברה הזאת וביטוח לאומי וצריך לעשות משהו
  3. בשורה שבה הלקוח(customer) הינו iaf נראה שהבונוס לסוף שבוע לא הגיוני – 500 אחוז בונוס על כל קילומטר. בנוסף הבונוס ללילה – 200 אחוז תוספת על כל קילומטר. - הגיוני
  4. בשורה שבה הלקוח(customer) הינו angel נראה שהבונוס לנסיעות ליליות יחסית גדול – 100 אחוז בונוס על כל קילומטר, האם הגיוני? אמר שהגיוני

חוסרים במידע:

* 1. בשורה שבה הלקוח (customer) הינו ביטוח לאומי (bituch\_leumi) השורה היא ללא מידע בכלל. – חסר וצריך לחשוב מה נעשה
  2. בשורה שבה הלקוח (customer) הינו electricity חסר מה התעריף הבסיסי (basic\_taarif). – חסר וצריך לחשוב מה נעשה

1. בקבצים **new\_drivers.csv, Drivers\_with\_kviut.csv** ישנו מידע חסר:
   1. אם המגדר חסר – אני כרגע ממלא Unknown.
   2. חלק מהתאריכי לידה חסרים בשני הקבצים הללו – כרגע ממלא לא ידוע.

שאלות נוספות:

1. מה ההבדל בין לנהג עם קביעות לבלי קביעות? האם זה מעל שנה פשוט? – לפי מה שהוא רשם אין הבדל סתם איך מודדים את הוותק
2. ישנם רשומות במידע על הנסיעות ללא זמן תחילת הנסיעה או סוף הנסיעה – מה לעשות במקרה זה? במצב כזה אני לא יודע כמה זמן לקחה הנסיעה ומתי היא התחילה והסתיימה. – לחשב ממוצע זמן נסיעה לקילומטר ולחשב זמן לפי זה
3. קיימות רשומות של לקוחות (customers) במידע על נסיעות שאינן נמצאות בקובץ taarif.csv. (לדוגמה aminach קיימת באחד מקבצי הנסיעות אבל איננה בקובץ taarif.csv) – שם שונה בקובץ פשוט (מר עזריאלי שלח במייל)
4. לגבי חוסרים בדאטה בבונוסים – למלא ב20 אחוז פשוט

שאלות נוספות נוספות:

1. חסרים תעריפים של חברת חשמל ושל ביטוח לאומי – מה נעשה?

נקודות לפגישה עם מר עזריאלי

1. מה עושים לגבי נסיעות שחסר שם זמנים? – רעיון חוץ ממוצע (כרגע הרעיון הוא לעשות זמן לפי ממוצע נסיעות של הנהג או משהו בסגנון ולמצוא את הזמנים שהוא נסע בהם)
2. ביטוח לאומי וחברת חשמל – הפסקתם לעבוד איתם? כן אין מידע וצריך להשלים לבד עם ממוצע נגיד
3. חברות שמשלמות 0 בונוס – לבדוק עם ברוריה – נבדק וכתבתי על זה – הגדלתי ל20 את כל הבונוסים שהיו 0
4. חברות שיש בקבצים של הנסיעות אך לא בקובץ התעריפים – עמינח, יס, הוט (aminah, yes, hot) – השלמנו יש להם פשוט שם אחר בקובץ תעריפים :

Aminach -> mizranei\_kfar\_saba  
Yes -> dbs\_satelites  
Hot -> telecommunication\_ltd

1. האם התשלום הנוסף לכל קילומטר (extra milage) אמור להיות תשלום בנוסף לתשלום בסיסי לכל קילומטר? אם לא אז יש בעיות כי חלק מהתשלומים על.
2. איך אני אמור לדעת כמה קילומטרים הוא נסע ב10 דקות האלה בסופש? לעשות ממוצע פשוט נכון? - כן
3. האם התשלום של סופש כולל את שעה 8 בערב עד 9 בערב? אותו דבר לתעריף לילה ו6 בבוקר – לא

עוד דברים:

1. לגבי חברת חשמל וביטוח לאומי – אין נתונים אז פשוט נעשה ממוצע של שאר החברות ונשלם לפי זה.
2. חיל האוויר משלם יותר מיד – מר עזריאלי ביקש להוריד 0 מהסוף מהבונוס לילה ובונוס סופ"ש שלהם.
3. ירשולים מוני משלם יותר מיד – מר עזריאלי ביקש להוריד 0 מהסוף מהבונוס לקילומטרים אקסטרה אצלם.
4. קבצים כפולים מוסרים בעיבוד מקדים
5. תיקון הקמ"ש שיתפלג נורמלי (יש outliers שהקמש שם מתחת לחמישים או מעל 100)– עוד לא סיימת וגם לא הבנתי כל כך מה עשיתי שם אבל זה עבד
6. תיקון השעות לפי פרדיקציה של הקמ"ש – עוד לא עשיתי (מה קורה כשחסר גם שעת התחלה וגם סיום?) (כרגע אם חסר סיום והתחלה פשוט אתחיל מ6 בבוקר בראשון כנראה)
7. אלה עם המון קילומטרים (4000 עד 8000 בערך) נזיז את הכמות קילומטרים שלהם לכמות הממוצעת ונתקן את זמן הסיום שלהם לפי המודל שלי.
8. אלה עם קצת קילומטרים נשאיר כמו שזה כי כנראה הם שכחו את זה דולק והפסיקו לנסוע ולכן רק נמצא עם המודל מה המהירות שלהם ונשים זמן סיום חדש.
9. תיקון לאלה עם שעת סיום או התחלה חסרים – פרדיקציה לקמש -> זמן נסיעה -> השלמת זמן התחלה\סיום.

פגישה 2 עם מר עזריאלי:

1. מה נעשה עם העניין שאין מידע לגבי חברות מסויימות לתעריפים? ממוצע של חברות אחרות
2. עדיין לא מרוצה מהרעיון של ממוצע של הבן אדם של הנסיעות למקרה אם חסר זמן התחלה או סיום? לשלוח ממוצעים
3. לגבי גילאים של אנשים – כרגע אני לדעתי אוריד לגמרי את עמודת הגיל פשוט

חסר ממש בהכשרה: איך לעבוד בפרוייקט יחסית גדול עם פנדה – מחברת\קלאסים. איזה דיזיין עושים ואיך שומרים על שמות המשתנים שיש בטבלאות וכו.....