

## תיעוד הפרויקט

### חישובים והגדרות:

1. קודם מחשבים את התשלום לקילומטר ועליו מוסיפים בונסים.
2. נסיעה בלילה ובסופש – מחשבים את הבנוס כסכום האחוזים של הבונסים בנפרד (לדוגמה אם נסענו בשבת בלילה והחברה מביאה לנו 30 אחוז בנוס לילה ו50 אחוז בנוס סופ"ש אז הבנוס לכל קילומטר בלילה בסופ"ש יהיה 80 אחוז.
3. שעות לילה – 22:00 – 6:00
4. שעות של סופ"ש – 16:00 – ביום שישי עד 20:00 בשבת.

### חוסרים והשלמות:

1. חברת חשמל וחברת ביטוח לאומי – אין מידע ולכן סימלצתי ממוצע של שאר החברות. (אצל חברת חשמל היה חלק מהמידע אבל הוא היה טעות)
2. חברות שמשלמות בנוס של 0 בלילות סופ"שים – ברוריה אמרה להעלות ל20 אחוז ישנם מספר חברות שהופיעו בנסיעות אך לא בקובץ התעריפים והם בעצם היו שם נרדף לחברה שכן בקובץ. החברות הן עמינח, יס והוט.
3. הורדתי 0 מהסוף בתשלום של הבונסים של חיל האוויר כי הוא שילם יותר מדי.

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

birthdate	gender	id	vetek	kviut
206	NaT	m	779	18.0
338	NaT	m	883	15.3
415	NaT	m	388	1.6
525	NaT	m	547	7.5
590	NaT	m	397	12.9
670	NaT	m	785	14.7
701	NaT	m	871	7.2
708	NaT	m	347	15.2
755	NaT	m	668	14.9
901	NaT	m	690	8.8

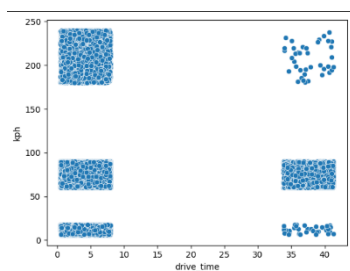
1. חוסר במגדר אצל נהג – משאיר כחסר מגדר
2. היה מגדר גבר ואישה כתוב בדרכים שונות ואיחדתי לפורמט אחד של m,f
3. חוסר בתאריך לידה – השארתי ללא גיל (בתמונה אלה ללא תאריך לידה)
4. קבצים כפולים בנסיעות - הסרתי
5. כפילויות בתוך קבצי הנסיעות של נסיעה כפולה - הסרתי

### תיקון שגיאות וחוסרים:

1. יצרתי מודל שבהינתן כמות קילומטרים ומידע על היום ושעה שבו נסע נהג מוציא זמן הגיוני שלקחה הנסיעה (בעצם מוציא קמש ממוצע שזה בגדול אותו דבר)
2. נסיעות ללא זמן התחלה וללא זמן סיום - שמתי את הזמן התחלה שלהם יום ראשון בשעה 6 בבוקר (מינימום כסף שאפשר לשלם להם לפי החישובי בונסים – כלומר יקבלו בונס רק אם נסעו עד אחרי 22 בלילה)
3. נסיעות ללא זמן התחלה או סוף – שימוש במודל נפרד לכאלה ללא זמן התחלה ולכאלה ללא זמן סיום אשר אומן על הדאטה ללא outliers לפי מבחן z עם threshold מסוים שנבחר לפי הסתכלות על איך המידע מתפלג בהינתן thresholds השונים. השימוש במודל נעשה על מנת לחזות כמה זמן נסעו בהינתן היום והשעה של הזמן התחלה/סיום.

### 4. תיקון outliers עם יותר מדי זמן נסיעה -

- אפשר לראות בתמונה – הגושי נקודות הדלילים בפינות של צד ימין.  
(א) את outliers תיקנתי באמצעות כך שאלה שנסעו נסיעות מאוד ארוכות (4000-8000 קילומטר), הקטנתי את הנסיעה שלהם לממוצע הנסיעות שלהם.  
(ב) לאחר מכן השתמשתי במודל על מנת לחשב כמות זמן הגיונית לפי הקילומטרים והזמן התחלה שהם נסעו בו.  
(ג) outliers שנסעו כמות קילומטרים הגיונית (outliers בצד ימין



למטה בתמונה) חוץ מהעניין של להקטין את הכמות קילומטרים שהם עשו מכיוון שכנראה הטעות שם זה ששכחו את המונה זמן דולק והפסיקו לנסוע, לכן הכמות קילומטרים כנראה תקינה.

#### **נוספים:**

1. קביעות – אין לזה משמעות מבחינת מר עזריאלי