**Preprocessing**

הספריה הנפוצה ביותר היא של Sklearn.

פוסט עם דוגמאות הקוד:

<https://towardsdatascience.com/preprocessing-with-sklearn-a-complete-and-comprehensive-guide-670cb98fcfb9>

טיפול בmissing values:

1. להשאיר ולתת לאלגוריתם לטפל במידה ויודע
2. להעיף רשומה בעזרת פונקציית dropna
3. להשלים ערך לפי מדד מסוים (לדוגמא ממוצע) fillna

הוספת משתנים פולינומיאלים (polynomial)

1. להוסיף משתנים עם Sklearn כמו הדוגמא בפוסט
2. להוסיף בדרכים אחרות כמו: <https://tsfresh.readthedocs.io/en/latest/>

משתנים קטגוריאלים:

1. להשאיר ולתת לאלגוריתם לטפל במידה ויודע
2. להמיר לערכים מספריים לדוגמא OrdinalEncoder
3. להפוך לdammy variable לדוגמא עם OneHotEncoder

\*\*דוגמאות בפוסט

משתנים מספריים:

1. להשאיר ולתת לאלגוריתם לטפל במידה ויודע
2. החלפה לbins.
3. להפוך לdammy variable לדוגמא עם OneHotEncoder

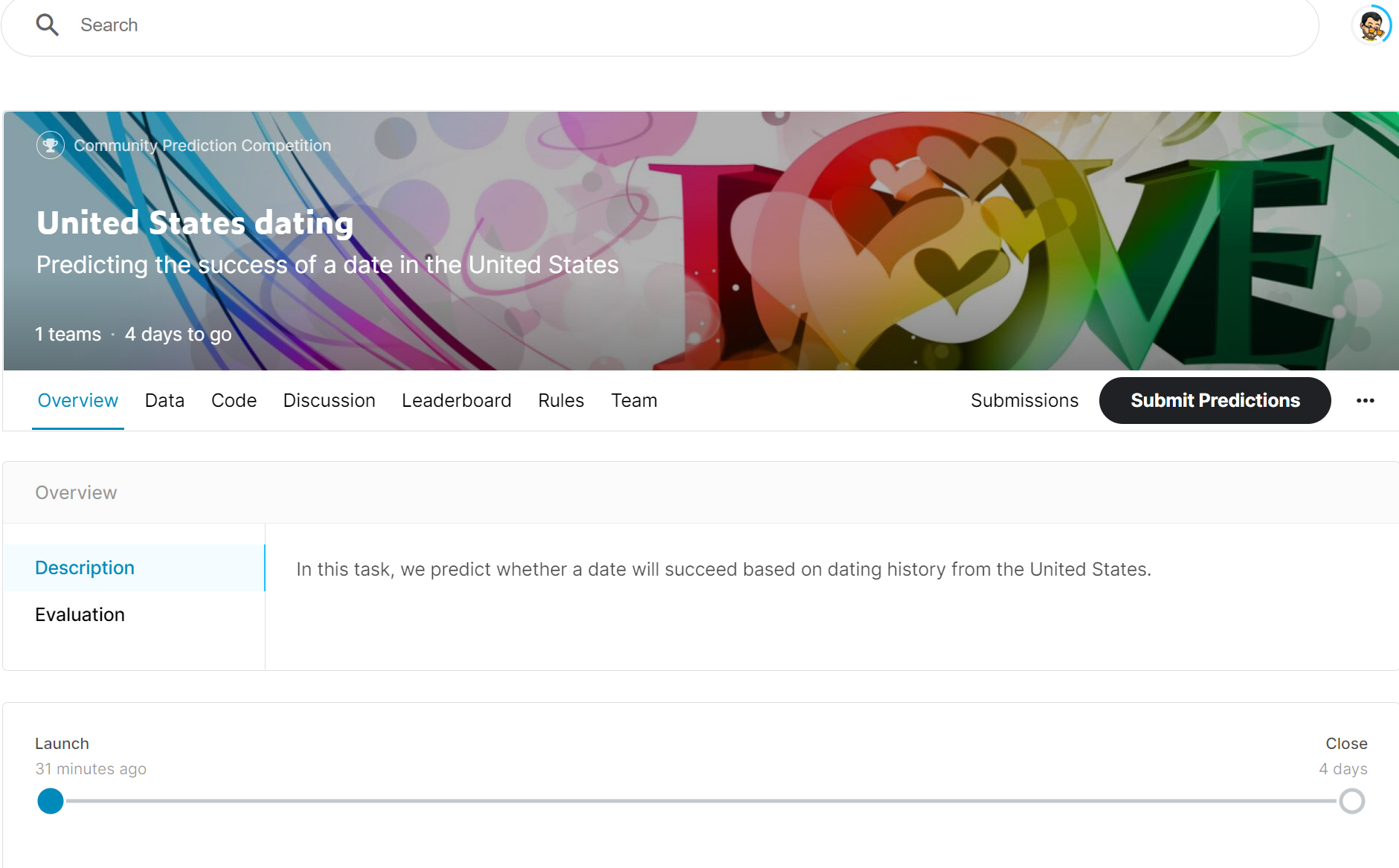
\*\*דוגמאות בפוסט

**תחרות KAGGLE**

לאחר יצירת משתמש בקאגל ניתן להתחבר לתחרות בעזרת הלינק:

<https://www.kaggle.com/t/4163149ad19242bdbc703f926a16d330>

תחרות נראית באופן הבא:



Overview – מכיל תיאור תחרות ושיטת הערכה (evaluation), בשיטת מצוינת מטריקת בדיקת התוצאות ומבנה קובץ ההגשה.

Data – קבצי המידע לתחרות, פירוט מאפיינים וסטטיסטיקה בסיסית עליהם.

Code – לשונית בה ניתן לראות את כל המחברות שמשתתפי התחרות שיתפו. כרגע יש שם מחברת אחת לדוגמא שניתן להשתמש בה. (אם תריצו את הקוד הזה בדיוק תקבלו את התוצאה של המשתמש שלי)

Discussion – מקום לדיונים בנוגע לתחרות

Leaderboard – טבלת ניקוד שמחולקת לציבורית ופרטית. טבלה ציבורית זה הציון שניתן עבור החלק בtest שהוא ציבורי (כלומר רואים את התוצאה שלו מיד בתום ההגשה). החלק הפרטי מפורסם אוטומטית כיממה לאחר סיום התחרות ולפיו נקבעים הדירוגים האמיתיים. (החלק הציבורי נותן הערכה לבדיקה האמיתית).

Rules & teams – חוקים וקבוצות. בתחרויות בקורס הקבוצות הן של 1.

Submittions – היסטוריית ההגשות שלי יחד עם הציון הציבורי שלהן. ניתן לבחור איזה מההגשות שלי ייבחנו בבדיקה הפרטית (בוחרים 2 הגשות). במידה ולא בוחרים ידנית, המערכת בוחרת אוטומטית את ההגשות שקיבלו את הציון הגבוה יותר בתוצאה הציבורית.

Submit predicitions – מקום ההגשה. שימו לב להגבלה כמות ההגשות וגם לפורמט קובץ ההגשה. (אפשר להריץ את המחברת שלי כדי לראות איך נראה קובץ הגשה).