

# Necmettin Erbakan Üniversitesi

## Veri Önişlemeye Giriş Ödev Raporu



Ahmet Furkan DEMİR  
19010011019

Danışman: MURAT KARAKOYUN

# San Francisco İnşaat İzinleri Veriseti

Veri seti Linki: [Link](#)



→ Veri hakkında:

İnşaat ruhsatı, sizin veya müteahhitinizin bir kişinin mülkü üzerinde bir inşaat veya tadilat projesine devam etmesine izin veren bir devlet kurumu tarafından verilen resmi bir onay belgesidir. Kullanacağımız bu veri seti San Francisco bölgesine ait inşaat ruhsatı verilerini kapsamaktadır.

→ Bu veri seti ile yapılabilecek bazı tahminler;

- Çeşitli izin türleri için izin verme sürelerini tahmin etmek.
- Denetim Departmanı'nı ziyaret etmek için en iyi günü tahmin etmek.
- Şehrin gelişmeleri hakkında bir sonuc tahmini

- ➔ Veri seti toplamda 198900 satırdan oluşmaktadır, verinin bulunduğu .csv dosyasının boyutu ise 75.4 MB. Veri seti sütunları 14 Integer, 11 String, 3 Float, 6 DateTime ve 9 NaN (Boş) veri tiplerine sahiptir.
- ➔ Veri setinde sınıflandırma olarak kullanılabilecek olan öznitelik Current Status (inşaatin şu an ki durumu) sütunudur. Bu sütun da 14 farklı sınıf bulunmaktadır ('expired', 'issued', 'withdrawn', 'complete', 'filed', 'approved', 'reinstated', 'cancelled', 'appeal', 'suspend', 'revoked', 'plancheck', 'disapproved', 'incomplete').
- ➔ Veri seti sütunları ve satır türleri (2/1):
- ✓ Satırların veri türü kısaltmaları  
S: String, I: Integer, F: Float, DT: DateTime, N: NaN (Boş)
- ◆ Permit Number <-> S: 37263, I: 0, F: 161637, DT: 0, N: 0
- ◆ Permit Type <-> S: 0, I: 198900, F: 0, DT: 0, N: 0
- ◆ Permit Type Definition <-> S: 198900, I: 0, F: 0, DT: 0, N: 0
- ◆ Permit Creation Date <-> S: 0, I: 0, F: 0, DT: 198900, N: 0
- ◆ Block <-> S: 7230, I: 0, F: 191670, DT: 0, N: 0
- ◆ Lot <-> S: 15875, I: 0, F: 183025, DT: 0, N: 0
- ◆ Street Number <-> S: 0, I: 198900, F: 0, DT: 0, N: 0
- ◆ Street Number Suffix <-> S: 2215, I: 0, F: 1, DT: 0, N: 196684
- ◆ Street Name <-> S: 198900, I: 0, F: 0, DT: 0, N: 0
- ◆ Street Suffix <-> S: 196132, I: 0, F: 0, DT: 0, N: 2768
- ◆ Unit <-> S: 0, I: 29479, F: 0, DT: 0, N: 169421
- ◆ Unit Suffix <-> S: 1951, I: 0, F: 10, DT: 0, N: 196939
- ◆ Description <-> S: 191863, I: 0, F: 4, DT: 6743, N: 290
- ◆ Current Status <-> S: 198900, I: 0, F: 0, DT: 0, N: 0
- ◆ Current Status Date <-> S: 0, I: 0, F: 0, DT: 198900, N: 0
- ◆ Filed Date <-> S: 0, I: 0, F: 0, DT: 198900, N: 0
- ◆ Issued Date <-> S: 0, I: 0, F: 0, DT: 183960, N: 14940

→ Veri seti sütunları ve satır türleri (2/2):

✓ Satırların veri türü kısaltmaları

S: String, I: Integer, F: Float, DT: DateTime, N: NaN (Boş)

- ◆ Completed Date <-> S: 0, I: 0, F: 0, DT: 97191, N: 101709
- ◆ First Construction Document Date <-> S: 0, I: 0, F: 0, DT: 183954, N: 14946
- ◆ Structural Notification <-> S: 6922, I: 0, F: 0, DT: 0, N: 191978
- ◆ Number of Existing Stories <-> S: 0, I: 156113, F: 3, DT: 0, N: 42784
- ◆ Number of Proposed Stories <-> S: 0, I: 156029, F: 3, DT: 0, N: 42868
- ◆ Voluntary Soft-Story Retrofit <-> S: 35, I: 0, F: 0, DT: 0, N: 198865
- ◆ Fire Only Permit <-> S: 18827, I: 0, F: 0, DT: 0, N: 180073
- ◆ Permit Expiration Date <-> S: 0, I: 0, F: 0, DT: 147020, N: 51880
- ◆ Estimated Cost <-> S: 0, I: 160247, F: 587, DT: 0, N: 38066
- ◆ Revised Cost <-> S: 0, I: 190535, F: 2299, DT: 0, N: 6066
- ◆ Existing Use <-> S: 157786, I: 0, F: 0, DT: 0, N: 41114
- ◆ Existing Units <-> S: 0, I: 147361, F: 1, DT: 0, N: 51538
- ◆ Proposed Use <-> S: 156461, I: 0, F: 0, DT: 0, N: 42439
- ◆ Proposed Units <-> S: 0, I: 147989, F: 0, DT: 0, N: 50911
- ◆ Plansets <-> S: 0, I: 161591, F: 0, DT: 0, N: 37309
- ◆ TIDF Compliance <-> S: 2, I: 0, F: 0, DT: 0, N: 198898
- ◆ Existing Construction Type <-> S: 0, I: 155534, F: 0, DT: 0, N: 43366
- ◆ Existing Construction Type Description <-> S: 155534, I: 0, F: 0, DT: 0, N: 43366
- ◆ Proposed Construction Type <-> S: 0, I: 155738, F: 0, DT: 0, N: 43162
- ◆ Proposed Construction Type Description <-> S: 155738, I: 0, F: 0, DT: 0, N: 43162
- ◆ Site Permit <-> S: 5359, I: 0, F: 0, DT: 0, N: 193541
- ◆ Supervisor District <-> S: 0, I: 197183, F: 0, DT: 0, N: 1717
- ◆ Neighborhoods - Analysis Boundaries <-> S: 194521, I: 0, F: 0, DT: 2654, N: 1725
- ◆ Zipcode <-> S: 0, I: 197184, F: 0, DT: 0, N: 1716
- ◆ Location <-> S: 197200, I: 0, F: 0, DT: 0, N: 1700
- ◆ Record ID <-> S: 0, I: 198900, F: 0, DT: 0, N: 0

## → Correlogram

Korelogram, belirli bir veri çerçevesindeki (veya 2B dizideki) tüm olası sayısal değişken çiftleri arasındaki korelasyon metriğini görsel olarak görmek için kullanılır. Bu sayede öznitelikler arasında ki ilişkiyi görebiliriz ve çıkarılacak olan özneliğin diğer özneliğe etkisini ölçebiliriz.

