

## Veri Bilimci Mulakat Projesi

Bu projede hedeflenen sizin Data Science alanında yetkinliklerinizi ölçmektir. Veri kaynağı olarak iki farklı dosya sizinle özel olarak paylaşılacaktır. Sizden çıktı olarak Jupyter Notebook beklenmektedir. Jupyter ortamına hakim değilseniz PDF çıktı üreten bir Python scripti hazırlayabilirsiniz. Proje 4 bölümden oluşmaktadır ve her bölüm için yeterli gördüğünüz kadar içerik üretmekte özgürsünüz. Bu bölümler adayları değerlendirmede bir formata ihtiyaç duyulması sebebiyle belirtilmiştir. Paylaşılan veri setiyle burada bahsedilmeyen fayda yaratacağını düşündüğünüz analizleri/görselleştirmeleri ekstra olarak yapmakta özgürsünüz.

### Önemli

1. Bu projenin çıktısı tamamıyla size aittir. APPS Bilgi Teknolojileri tarafından ticari olarak kullanılması mümkün değildir.
2. Veri setindeki bilgiler anonimleştirilmiştir ve size özgün hazırlanmıştır. Dolayısıyla kamuyla paylaşılmamalıdır.
3. Projede mutlaka git kullanılmalıdır ve csv dosyaları ignore edilmelidir. APPS Bilgi Teknolojileri olarak projenizi daha kolay takip etmek açısından kendi git sunucumuzda barındırmayı tercih ederiz. Mantıklı sebeplerden projeyi siz host etmek isterseniz mutlaka private bir repository'ide olmalı ve mülakat süresince irtibatta olduğunuz kişilere gözlem yetkisi vermelisiniz.
4. Geliştirme esnasında her türlü konuda çekinmeden soru sorabilirsiniz. Değerlendirmemize olumsuz etkisi olması söz konusu değildir.

### 1. Exploratory Data Analysis

Bu bölümde sizden veri setini istatistik bilginizle irdelemenizi ve bilgi birikiminize bağlı olarak görselleştirme ve analiz yapmanızı istiyoruz. Örneğin: Level başına yapılan yanlış sayısının bar grafiği gibi. Veri setini inceledikten sonra gördüğünüz problemleri açıklayabilir veya görselleştirebilirsiniz. Ayrıca bu bölümde bir sonraki bölümlerde işinize yarayacak değişkenleri yaratabilirsiniz.

## 2. Retention Calculation

Kullanıcıların 7 gün boyunca içinde kalma oranlarını oyuna dahil oldukları günü referans olarak gösteriniz. Aşağıdaki gibi bir görsel oluşturmanız bekleniyor. Eğer örnek görseli oluşturamazsanız ağırlıklı ortalamaları normal çizgi grafik ile gösterebilirsiniz.

Cohort of App Launched > App Launched									
Users who came back on →									
Cohort ↓	New Users	Day 0	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
All	134,529	100.0%	26.1%	20.0%	14.8%	11.5%	9.4%	8.8%	8.1%
Aug 30	32,961	100.0%	31.3%	26.8%	20.2%	13.4%	10.2%	9.7%	8.1%
Aug 31	20,144	100.0%	21.6%	17.2%	13.7%	9.8%	8.4%	7.9%	
Sep 01	17,772	100.0%	22.0%	19.8%	14.2%	11.2%	9.6%		
Sep 02	21,455	100.0%	27.4%	21.3%	15.8%	11.7%			
Sep 03	18,566	100.0%	24.5%	16.4%	10.3%				
Sep 04	10,656	100.0%	29.2%	18.6%					
Sep 05	12,975	100.0%	26.8%						
Sep 06	5,651	100.0%							

## 3. RFM Analysis

[RFM \(Recency Frequency Monetary\) analizi](#) ile kullanıcı davranışlarını basit anlamda kategorilendirebiliriz. Bu model normalde kullanıcının satın alma yakınlığı, sıklığı ve miktarını kullanarak segmentler oluşturup çeşitli senaryolar geliştirilmesine olanak sağlıyor. Bu modele bağlı olarak sınıflandırma yaparken quartile yada k-means kullanabilirsiniz. Model için;

Recency olarak kullanıcının aktif gün sayısı,

Frequency olarak bir kullanıcının aktif gün bazında 14 level tamamlaması bir frekans değeri (bir kullanıcı bir aktif günde 22 level tamamlamışsa Frekansı 1, 38 level tamamlamışsa Frekansı 2 gibi),

Monetary olarak kullanıcının sağladığı toplam kazancı kullanmalısınız.

## 4. Churn Prediction

Paylaşılan veri setinden yeni değişkenler (feature) üreterek bir churn prediction modeli oluşturmanız beklenmektedir. Kullanıcıların 7. gün oyunda kalıp kalmayacaklarını tahmin eden bir model oluşturmalısınız. Modelin performansından daha çok (hızlı olması yada doğruluğunun %70 olması gibi) sizin probleme olan yaklaşımınız değerlendirilecektir. Herhangi bir model yada kütüphane sınırlaması yoktur.