

## GİRİŞ

Bu projede “Flight Delay and Causes” isimli veri seti temel alınarak, bu veri setinden elde edilen 2019 yılının 6 aylık ABD uçuş verileri ile Power BI üzerinde bir veri analizi ve veri görselleştirme çalışması yapılmıştır. Veri seti ve Power BI raporları aşağıda ayrıntılı olarak anlatılmıştır.

### 1. VERİ SETİ

“Flight Delay and Causes” veri seti Kaggle üzerinden alınmış olup, 2019 yılı Ocak-Haziran ayları arasındaki 6 aylık dilimde ABD içerisinde gerçekleşen uçuşlarla ilgili, tarih, havayolu şirketi, havaalanları ve IATA kodları ve çeşitli gecikme sebepleri gibi birçok veriyi içermektedir.

Veri Setinin Kaynağı: [www.kaggle.com/underscore/flight-delay-and-causes](https://www.kaggle.com/underscore/flight-delay-and-causes)

Veri setinde bulunan ve çalışmaya dahil edilen veri sütunları şunlardır;

**Date:** Planlanmış uçuş tarihi,

**DepTime:** Gerçek kalkış zamanı,

**ArrTime:** Gerçek varış zamanı,

**CRSArrTime:** Planlanan varış zamanı,

**UniqueCarrier:** Benzersiz taşıyıcı (araç, uçak) kodu,

**Airline:** Havayolu şirketi,

**FlightNum:** Uçuş numarası,

**TailNum:** Uçak kuyruk numarası,

**ActualElapsedTime:** Uçağın iniş ve kalkış süresi de dahil olmak üzere havada harcadığı süre (dakika),

**CRSElapsedTime:** Uçuşta harcanması planlanan süre (dakika),

**AirTime:** Havada geçen süre (dakika),

**ArrDelay:** Planlanan ve gerçek varış zamanı arasındaki dakika farkı,

**Origin:** Kalkış havaalanının 3 harfli IATA kodu,

**Org\_Airport:** Kalkış havaalanının adı,

**Dest:** Varış havaalanının 3 harfli IATA kodu,

**Dest\_Airport:** Varış havaalanının adı,

**Distance:** Kalkış ve varış havaalanları arasındaki mesafe (mil),

**TaxiIn:** Tekerleklerin indirilmesinden ve havaalanı kapısına varış anına kadar geçen dakika,

**TaxiOut:** Havaalanı kapısından hareket edilmesinden tekerleklerin kaldırılmasına kadar geçen dakika,

**CarrierDelay:** Havayolu şirketi nedeniyle (bakım, ekip sorunları, uçak temizliği, yakıt ikmali vb.) oluşan gecikme süresi (dakika),

**WeatherDelay:** Hava durumu nedeniyle uçuş gecikme süresi (dakika),

**NASDelay:** NAS (National Aviation System) tarafından oluşan uçuş gecikme süresi (dakika),

**SecurityDelay:** Güvenlik nedeniyle oluşan uçuş gecikme süresi (dakika),

**LateAircraft:** Geç ulaşan uçak nedeniyle oluşan gecikme süresi (dakika).

Bunların dışında veri setine çalışma dahilinde sonradan eklen veri sütunları şunlardır:

**WeekDay:** Uçuşun gerçekleştiği haftanın günü

**WeekDayNumber:** Haftanın günlerine göre uçuşun gerçekleştiği günün numara değeri (Pazartesi=1, Pazar=7)

Çalışmada “Flight Delay and Causes” veri setine ek olarak, harita üzerindeki veri görselleştirmelerinde kullanılmak amacıyla, OpenFlights.org - “OpenFlights Airports Database” üzerinden alınan uluslararası düzeyde 10000’den fazla havaalanı, feribot terminalleri ve tren istasyonlarını kapsayan bir veri seti daha kullanılmıştır. Bu veri seti içerisinde sadece “Airport Database” alınmış olup, çalışma için uygun sütunlar seçilmiştir. Veriler filtrelenerek sadece ABD üzerinde bulunan havaalanları çalışmaya dahil edilmiştir.

Veri Setinin Kaynağı: [openflights.org/data.html](https://openflights.org/data.html)

Veri setinde bulunan ve çalışmaya dahil edilen veri sütunları şunlardır;

**Name:** Havaalanının adı,

**City:** Havaalanının bulunduğu şehir,

**Country:** Havaalanının bulunduğu ülke veya bölge,

**IATA:** Havaalanının 3 harfli IATA kodu,

**Latitude:** Havaalanının enlem değeri,

**Longitude:** Havaalanının boylam değeri.

## 2. VERİ SETİ İSTATİSTİKLERİ

“Flight Delay and Causes” veri seti uçuşlardaki gecikme türlerine ve sürelerine odaklanmaktadır. Aşağıda veri setindeki bu gecikme türleri ve sürelerinin verildiği sütunların ortalama (average), mod (mode), medyan (median), basıklık (kurtosis), standart sapma (standart deviation) gibi bazı istatistiksel değerleri verilmiştir.

### 2.1. Gecikme Türleri Alanları

Aşağıdaki tablolarda veri setindeki gecikme türleri ve süreleri alanlarının istatistiki değerleri verilmiştir.

**Tablo 1.** Gecikme Türleri Alanlarının İstatistiksel Değerleri

	ArrDelay	DepDelay	CarrierDelay	Weather Delay	NASDelay	Security Delay	LateAircraft Delay
<b>Ortalama</b>	60,79	57,38	17,20,	3,15	13,65	0,08	26,7
<b>Mod</b>	15	25	0	0	0	0	0
<b>Medyan</b>	42	40	2	0	1	0	13
<b>Basıklık</b>	34,05621	36,0662	109,1286	285,2076	59,8380	5753,248	15,4653
<b>Standart Sapma</b>	56,77231	55,7602	39,0169	19,5018	31,5539	1,8793	40,4709

### 2.2. Diğer Sayısal Alanlar

Aşağıdaki tablolarda veri setindeki diğer sayısal alanlarının istatistiki değerleri verilmiştir.

**Tablo 2.** Diğer Sayısal Alanlarının İstatistiksel Değerleri

	ActualElapsedTime	CRSElapsedTime	AirTime	TaxiIn	TaxiOut
<b>Ortalama</b>	134,41	131	108,49	6,78	19,15
<b>Mod</b>	75	65	45	4	11
<b>Medyan</b>	116	113	89	5	15
<b>Basıklık</b>	2,3435	2,3813	2,5080	76,0584	31,9833
<b>Standart Sapma</b>	73,8435	71,3038	69,8824	5,5554	15,3482

### 3. POWER BI RAPORU

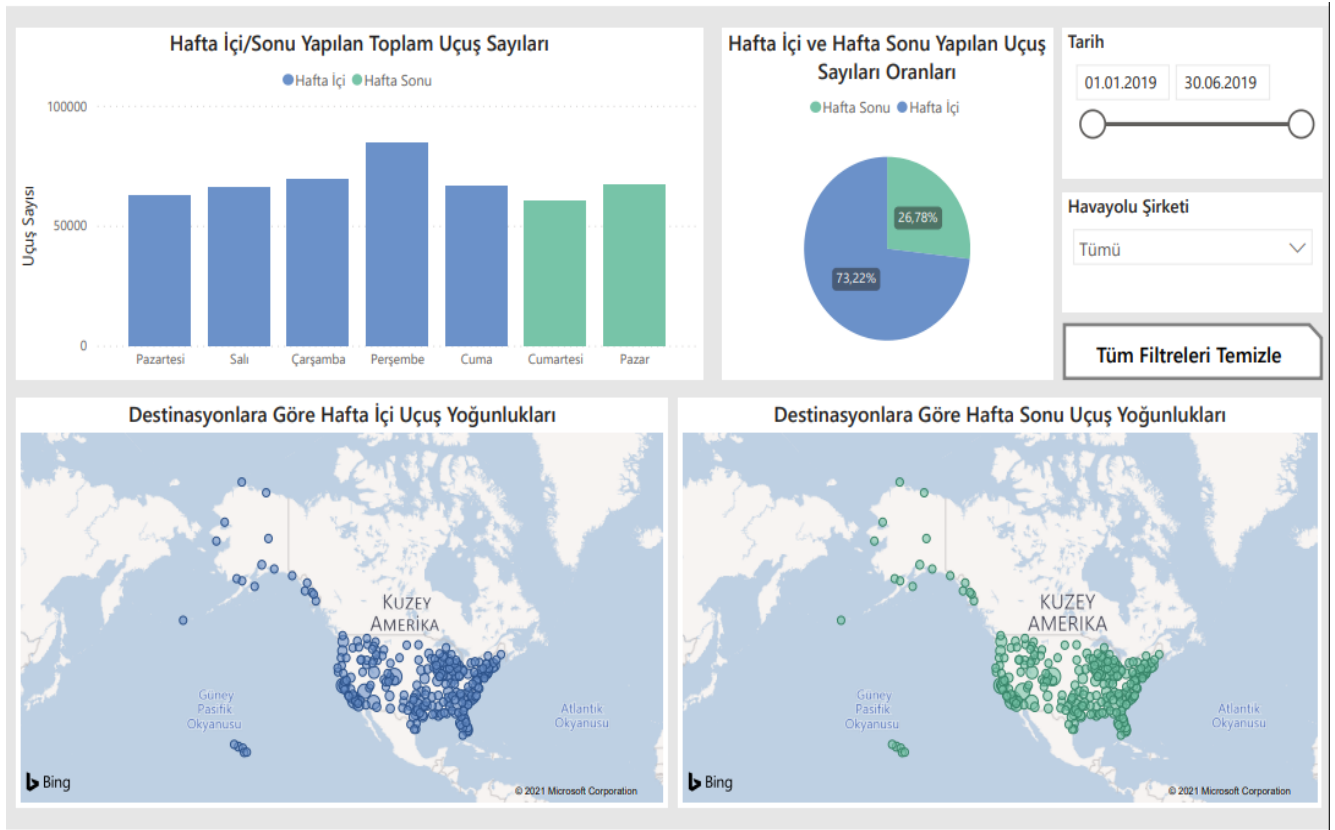
Veri setleri düzenlendikten sonra Power BI üzerinde veri görselleştirme çalışması yapılmıştır.

#### 3.1. Hafta İçi/Sonu Uçuş Yoğunlukları

Hafta İçi/Sonu Uçuş Yoğunlukları sayfasında veri setinin kapsadığı 6 aylık dilimde hafta içi ve hafta sonu yapılan toplam uçuş sayıları haftanın günlerine göre sütun grafiği üzerinde gösterilmiştir. Toplam hafta içi ve hafta sonu uçuş sayılarının birbirlerine oranları ise pasta grafiği üzerinde gösterilmiştir.

Hafta içi ve hafta sonu yapılan uçuşların destinasyonlara göre yoğunlukları harita üzerinde gösterilmiştir. Hangi destinasyonlara haftanın hangi bölümünde daha çok uçuş yapıldığı bu bölümden incelenebilmektedir.

Rapora eklenen dilimleyiciler (slicer) ile bu grafikler, tarih ve havayolu şirketleri bakımından filtrelenerek görüntülenebilir.



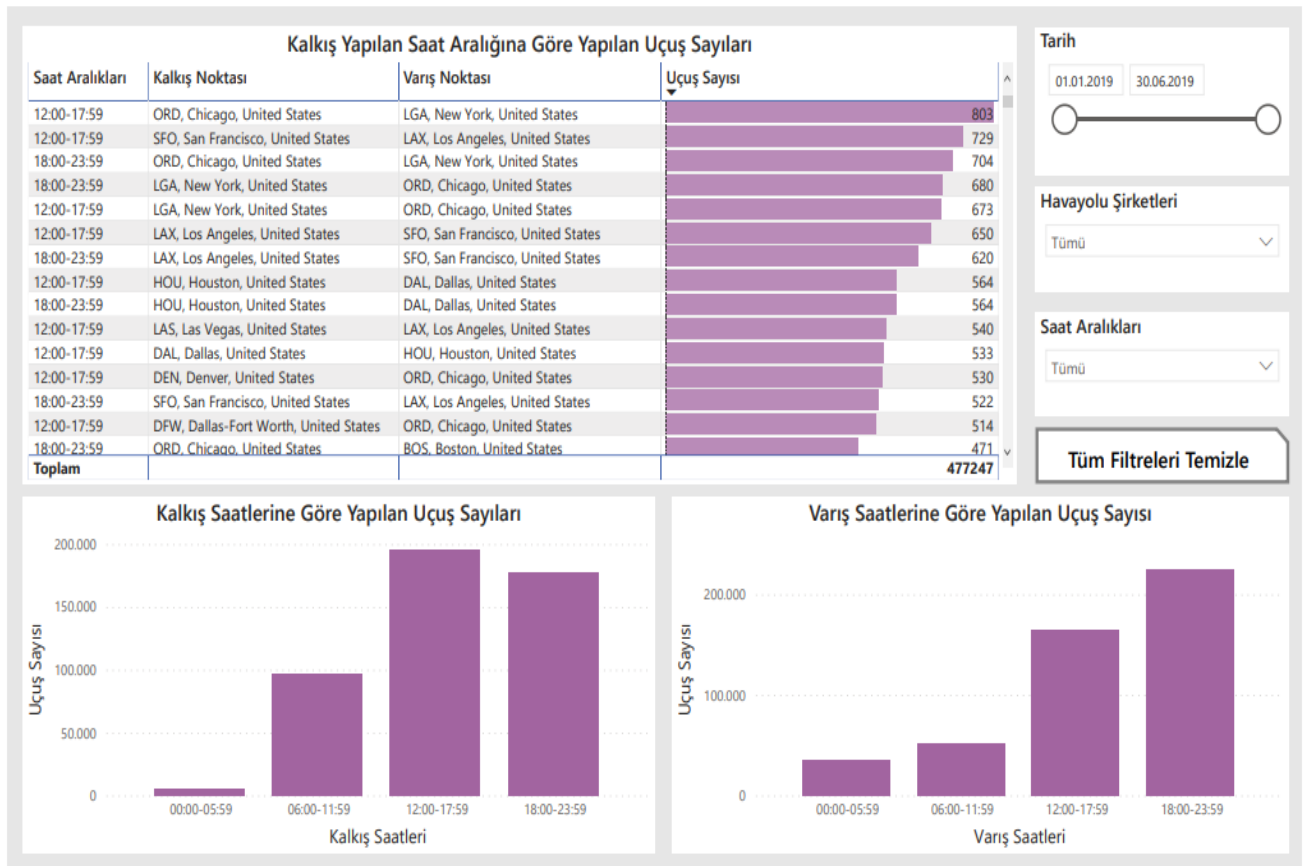
Şekil 1. Hafta İçi/Sonu Uçuş Yoğunlukları Sayfası

### 3.2. Saat Dilimleri Uçuş Yoğunlukları

Bu sayfada, 6 aylık tüm uçuş verilerinin kalkış ve varış saatlerinin 00:00-05:59, 06:00-11:59, 12:00-17:59, 18:00-23:59 olmak üzere 4 saat aralığında gruplandırılmış şekilde analizi yapılmıştır.

Uçuşların kalkış saatinin bulunduğu saat diliminde, kalkış noktası, varış noktası ve bu rota üzerindeki uçuş sayısı listelenerek tablo üzerinde gösterilmiştir. Bu tablo ile hangi rotada hangi saat aralığında daha fazla uçuşun yapıldığı, bu rotanın hangi saat diliminde daha yoğun olduğu hakkında bilgi edinilmesi amaçlanmıştır. Ayrıca sadece bu tabloya bağlı olarak eklenen dilimleyici ile saat aralığı filtresiyle listeleme yapılabilmektedir.

Sayfanın alt kısmında ise sütun grafikleri ile uçuşların kalkış saatinin bulunduğu saat aralığına göre uçuş sayıları ve varış saatinin bulunduğu saat aralığına göre uçuş sayıları verilmiş olup, hangi saat aralığının kalkış ve varış için daha yoğun olduğu gözlemlenebilmektedir.



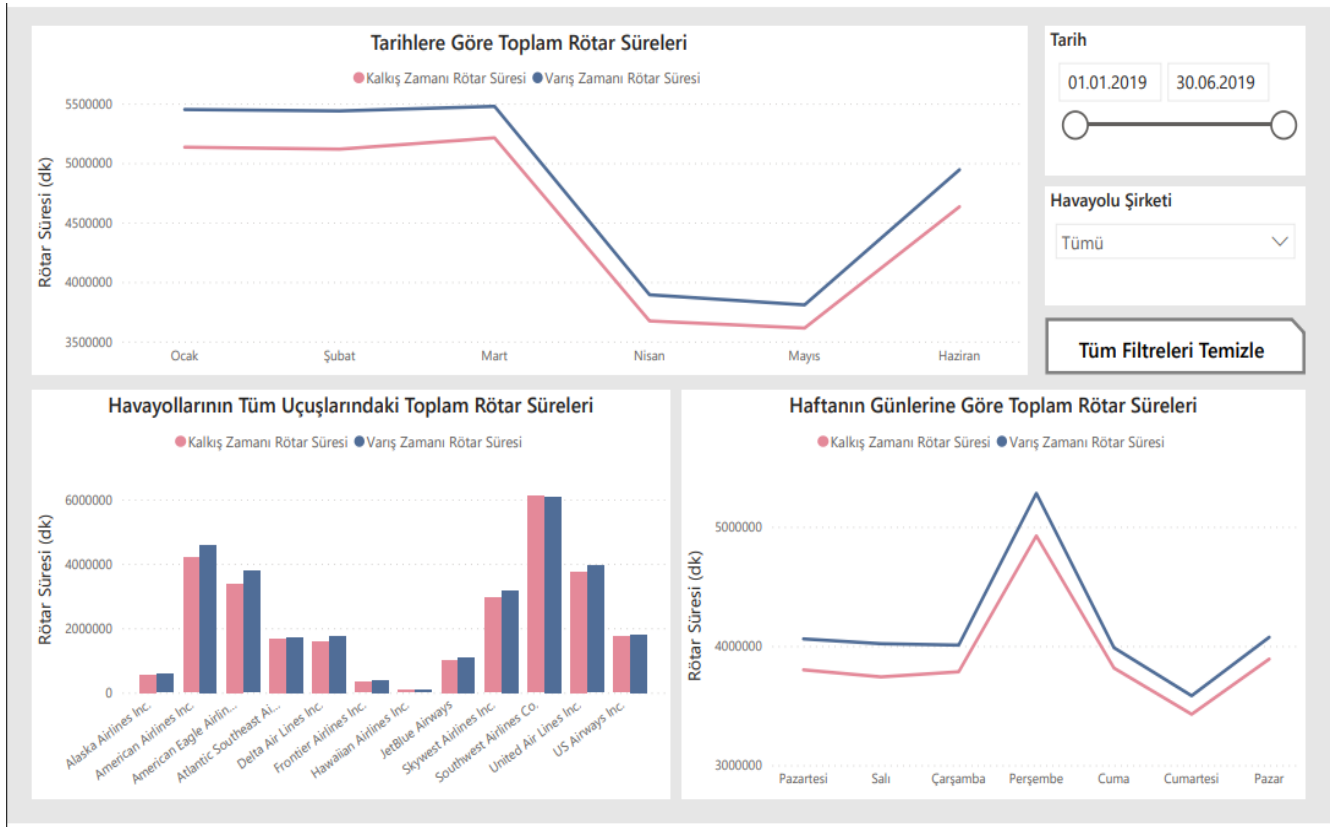
Şekil 2. Saat Dilimleri Uçuş Yoğunlukları Sayfası

### 3.3. Toplam Rötar Süreleri Analizi

Raporun bu sayfasında farklı ölçütlere göre toplam rötar süreleri incelenmiştir.

“Tarihlere Göre Toplam Rötar Süreleri” başlıklı çizgi grafiğinde, 6 aylık dilimde, her ay için kalkış zamanındaki rötar süreleri ve planlanan varış zamanına göre oluşan varış zamanı rötar süreleri gösterilmiştir. Benzer şekilde “Haftanın Günlerine Göre Toplam Rötar Süreleri” başlıklı çizgi grafiğinde de aynı alanlar için haftanın günlerine göre toplam rötar süreleri gösterilmiştir. Bu grafiklerle hangi aylarda ve haftanın hangi günlerinde daha çok rötar yaşandığı, ne kadar rötar yaşandığı anlaşılmaktadır. Aynı zamanda kalkış zamanında ortaya çıkan rötar süreleri ile varış zamanındaki rötar süreleri arasındaki ilişki de gözlenebilmektedir.

Bir diğer grafikte ise havayolu şirketlerine göre şirketlerin tüm uçuşlarındaki toplam rötar sürelerini gösteren bir grafik eklenmiştir.



Şekil 3. Toplam Rötar Süreleri Analiz Sayfası

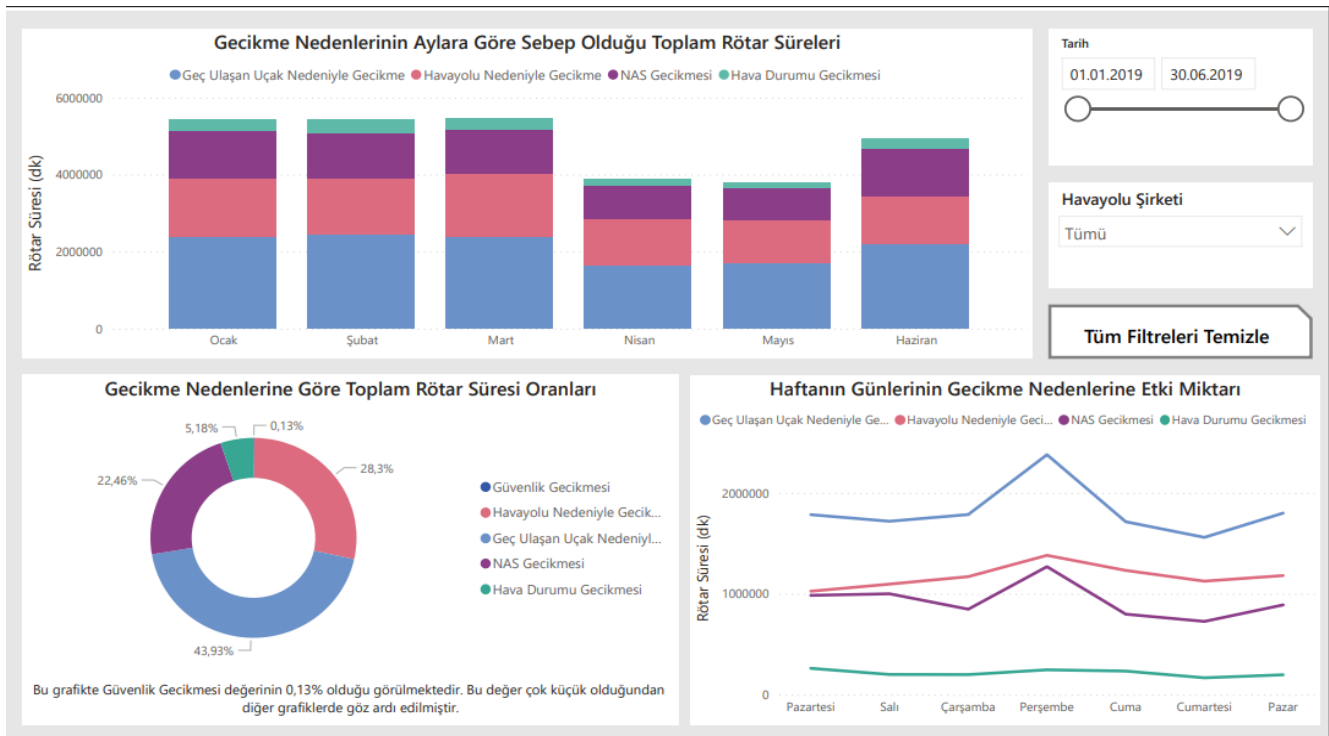
### 3.4.Rötar Nedenleri ve Süreleri

Veri setinde en çok rötara yol açan nedenlere ve bunların her uçuş için kaç dakika rötara sebep olduğuna yer verilmiştir. Bu sayfada da uçuşlar bu açıdan incelenmiştir.

Raporda veri setindeki verilere göre halka grafiği üzerinde beş rötar sebebinin neden olduğu toplam rötar sürelerinin birbirlerine oranları yüzdelik olarak verilmiştir. Bu grafikte bulunan “Güvenlik Gecikmesi” diğer nedenlere oranla %0,13’lük çok küçük bir paya sahiptir. Dolayısıyla bu rötar nedeni diğer grafiklerdeki değerlere katılmamıştır. 6 aylık süreçteki uçuşlarda gerçekleşen rötarların neden ve toplam kaç dakika olduğu ise sütun grafiği ile gösterilmiştir. Bu gösterimlerden yola çıkarak hangi rötara neyin daha sık sebep olduğu ve ne kadar fazla sürdüğü gözlenebilmektedir.

Ek olarak haftanın günlerine göre rötar sürelerinin dağılımını gösteren çizgi grafiği ile haftanın hangi günlerinde ne kadar rötar yaşandığının takip edilebilmesi sağlanmıştır.

Şekil 4’te sayfanın ekran görüntüsü verilmiştir, grafiklerde en yüksek değere “Geç Ulaşan Uçak Nedeniyle Gecikme”nin sahip olduğu görülmektedir. Bu bağlamda sonraki sayfada bu rötar nedeni ayrıntılı olarak incelenmiştir.



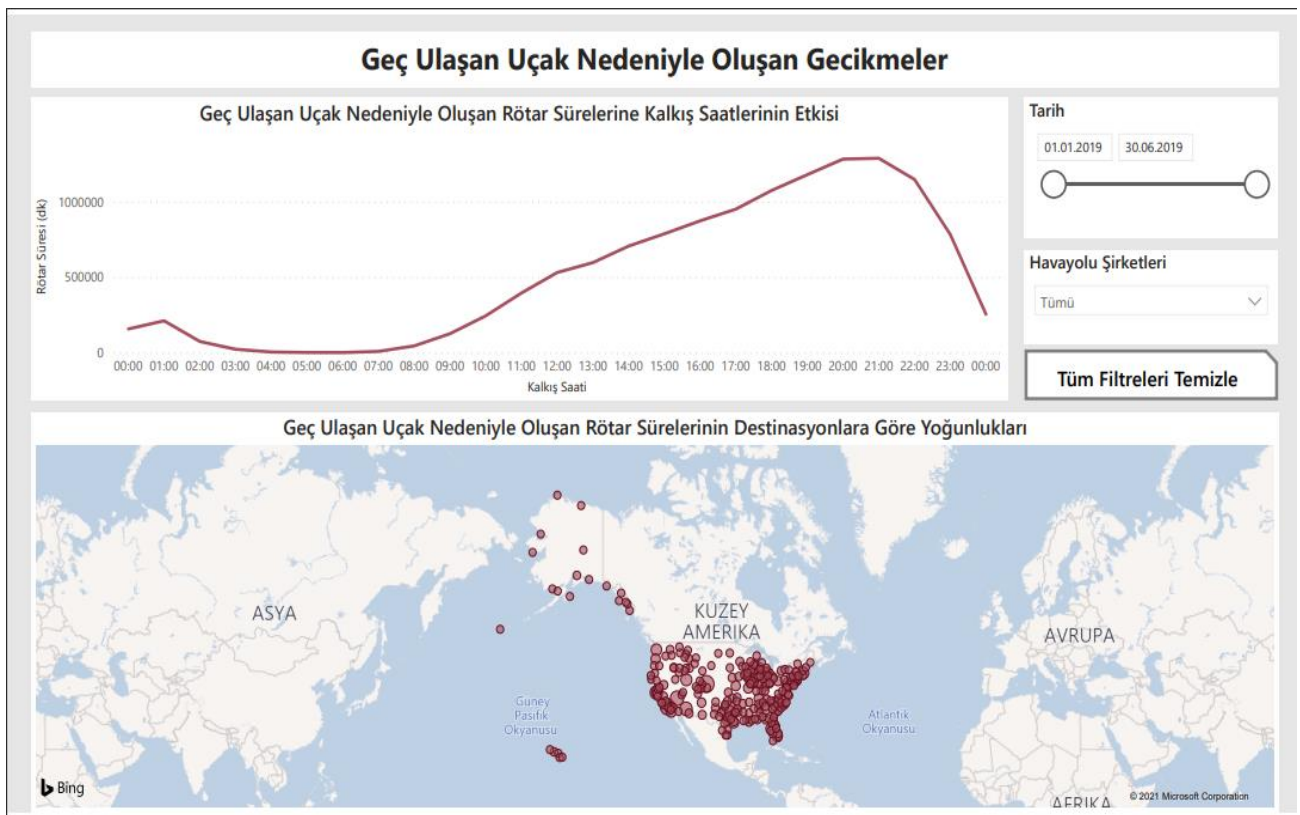
Şekil 4. Rötar Nedenleri ve Süreleri Analiz Sayfası

### 3.6. Geç Ulaşan Uçak Nedeniyle Gecikme Analizi

Daha önce yapılan analizlerden “Geç Ulaşan Uçak Nedeniyle Gecikme”nin yüksek değerlere sahip olduğu görülmüştür. Bu kısımda bu neden üzerine analizler yapılmıştır.

Çizgi grafiği üzerinde kalkış saatlerine göre gerçekleşen toplam rötar süreleri verilmiştir. Bu şekilde hangi saatlerde daha fazla rötar olduğu gözlenebilmektedir. Harita üzerinde de hangi destinasyonlara yapılan uçuşlarda daha yoğun rötar olduğu gösterilmiştir.

Eklenen dilimleyicilerle tarih ve havayolu şirketlerini filtreleyerek daha ayrıntılı inceleme yapılabilmektedir.



Şekil 5. Geç Ulaşan Uçak Nedeniyle Gecikme Analizi