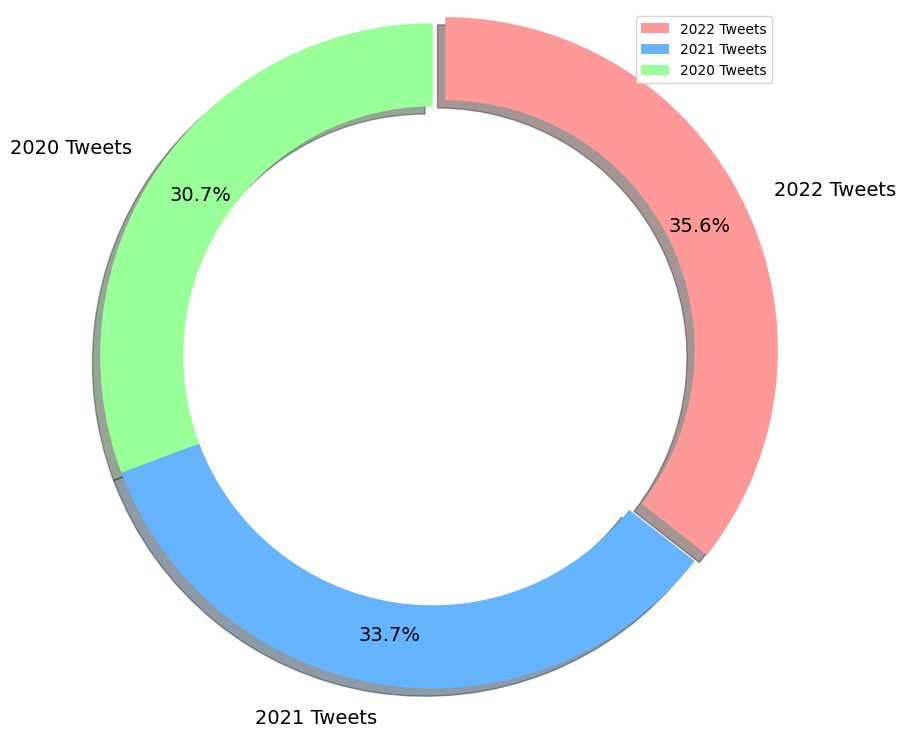
**2. Gereç ve Yöntem**

Bu araştırma, hasta güvenliği kavramının sosyal medya nasıl kullanıldığının belirlenmesi, kavramların yıllara ve lokasyonlara göre nasıl evrildiğinin ortaya konulması ile karar vericilere bir perspektif sunulması için tasarlanmıştır.

Araştırma, bir programlama dili olan Python (KAYNAK) kullanılarak gerçekleştirilmiştir. Bu bağlamda Twitter’dan patient safety anahtar kelimesi kullanılarak web scrapping (KAYNAK) yöntemi ile tweetler elde edilmiştir. Veri setinin analiz edilerek anlamlandırılmasında Topic Modeling (KAYNAK) yöntemi kullanılmıştır.

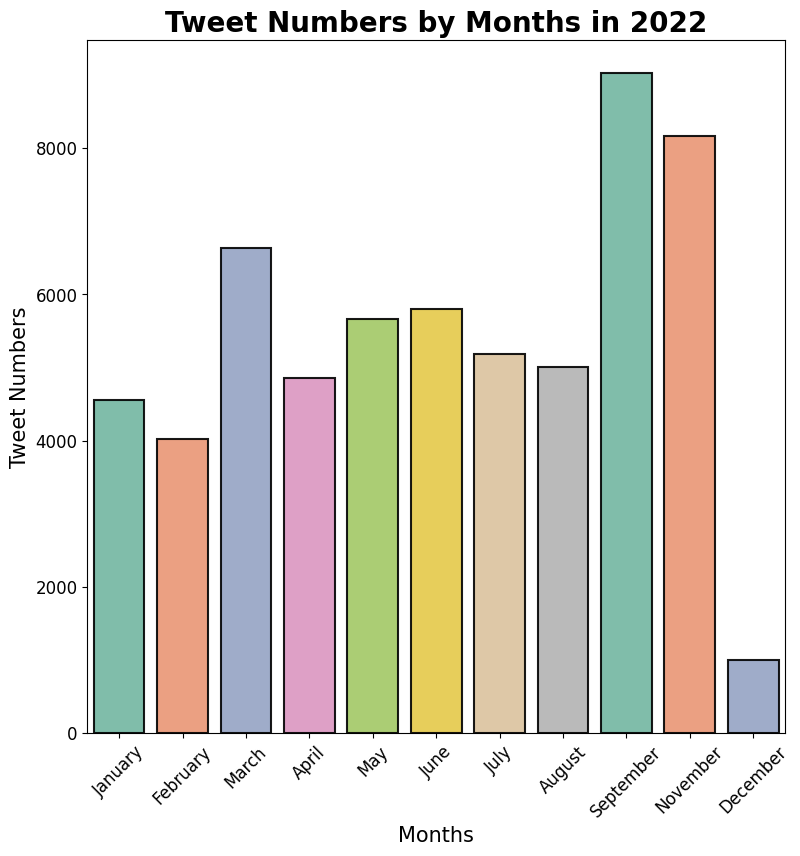
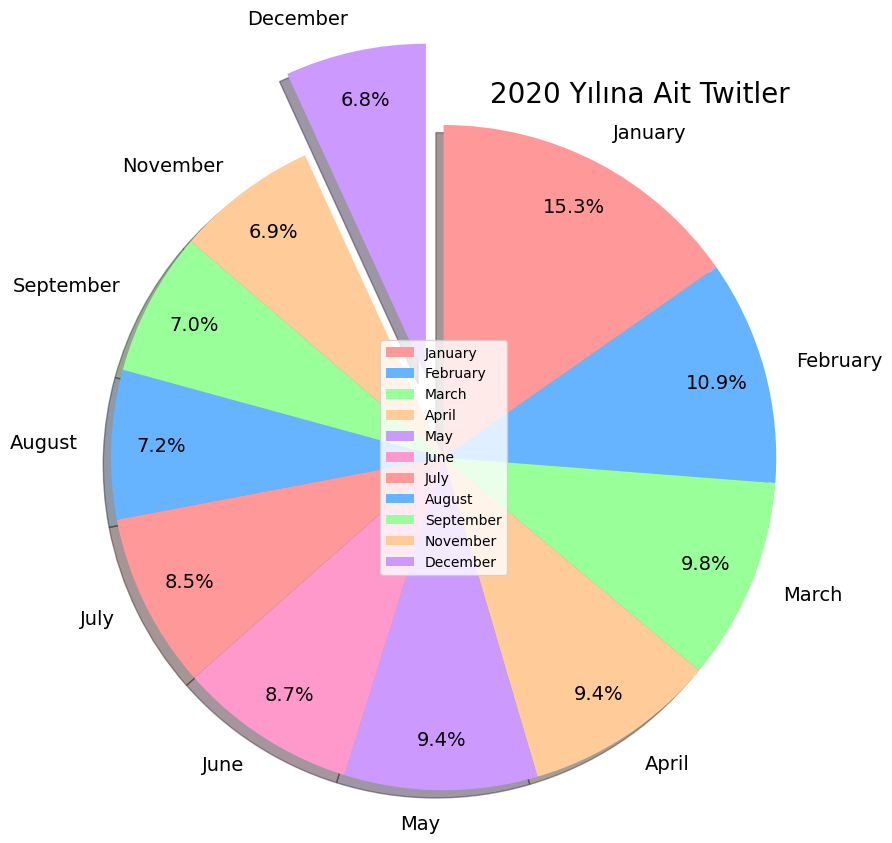
**2.1. Tanımlayıcı Bulgular**

Elde edilen 2020, 2021 ve 2022 yıllarına ait toplam 249.979 adet twitten oluşan hasta güvenliğine ilişkin veri seti analiz edilmiştir. Veri setinde 3509 adet duplikasyon twit bulunduğu belirlenmiş ve bunlar verisetinden çıkarılmıştır. Veri setinde 40 farklı dilden ve 53 farklı lokasyondan twit atıldığı belirlenmiştir. Araştırma, sonuç olarak 71.063 benzersiz kullanıcının atmış olduğu toplam 236.955 “İngilizce” twit dikkate alınarak gerçekleştirilmiştir. Yıllara sari twit sayıları Şekil 1’de gösterilmiştir.

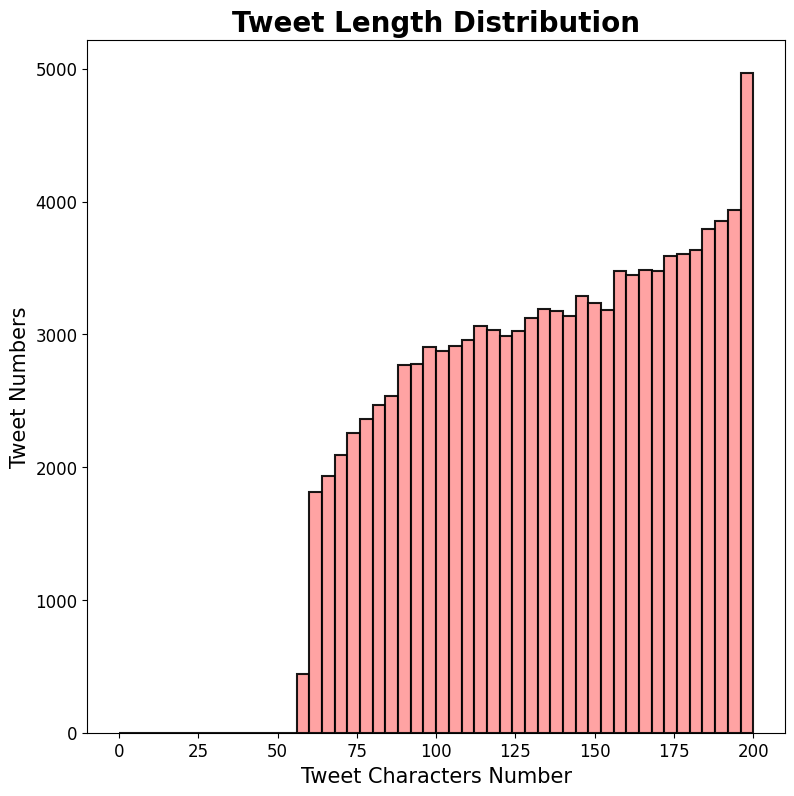


Buna göre atılan twit sayılarının 2020 yılında 71.487 (%30,7), 2021 yılında 61.715 (%33,7) ve 2022 yılında 67.728 (%35,6) olduğu belirlenmiştir.

Aylar bazında atılan twitler değerlendirildiğinde en fazla Ocak aylarında ve en az Eylül aylarında atıldığı görülmüştür. Örnektwit dağılımı oranları 2022 yılı için Şekil Şekil 2’de gösterilmiştir.

Elde edilen tweetler öncelikle bir temizleme işleminden geçirilmiş, metin semboller, fotoğraflar, piktoğraf, bayrak, harita resimleri, url linkleri vb. karakterlerden temizlenmiştir. Metin içerinden bulunan ve ‘#’ ve ‘@’ sembolleri ile başlayan, iki harf ve daha az harften oluşan kelimeler metinden çıkarılmıştır. Metin içinde herhangi boş bir değer olmadığı görülmüştür. Bütün bu temizleme işlemlerinin ardından toplam 200.930 adet twit ile işlemlere devam edilmiştir.

Atılan twitler tokenizer işlemine tabi tutulmuş ve twitlerin karakter sayıları analiz edilmiştir. Uç değerler belirlenmiş, 58 ve daha az karakterden oluşan twitler araştırmadan çıkarılmıştır (Şekil 3).

Yapay zekâ alt bileşenlerinden olan doğal dil işleme (Natural language processing, NLP)’de sıklıkla kullanılan stopwords ve lemmatizer (KAYNAK) kullanılarak veri setindeki gereksiz kelimeler temizlenmiş ve çekimli kelimelerin indirgeme işlemi yapılmıştır. Elde edilen kelime bulutu Şekil 5’te gösterilmiştir.



**2.2. Topic Model**