**BÖLÜM 6**

**Veri Yükleme, Depolama ve Dosya Biçimleri**

Verilere erişmek, araçların çoğunu kullanmak için gerekli ilk adımdır. Bu bölümde Pandas kullanarak veri girişine ve çıkışına bakacağız. Giriş ve çıkış tipik olarak birkaç ana kategoriye ayrılır: metin dosyalarını ve diğer daha verimli disk formatlarını okumak, veri tabanlarından veri yüklemek ve web API gibi ağ kaynaklarıyla etkileşim kurmak.

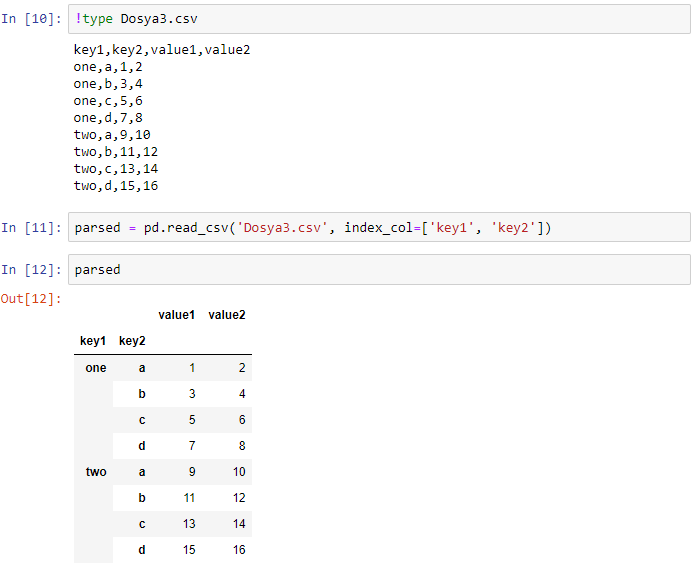
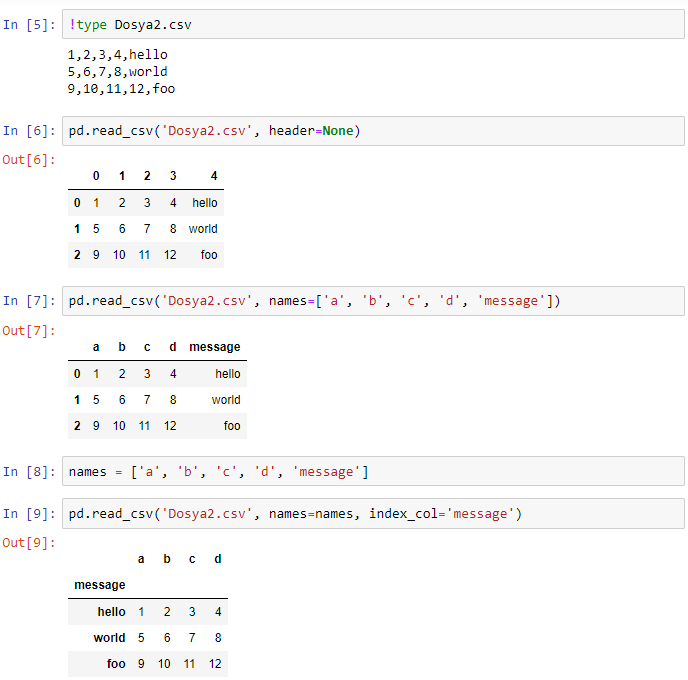
**6.1. Metin Formatında Veri Okuma ve Yazma**

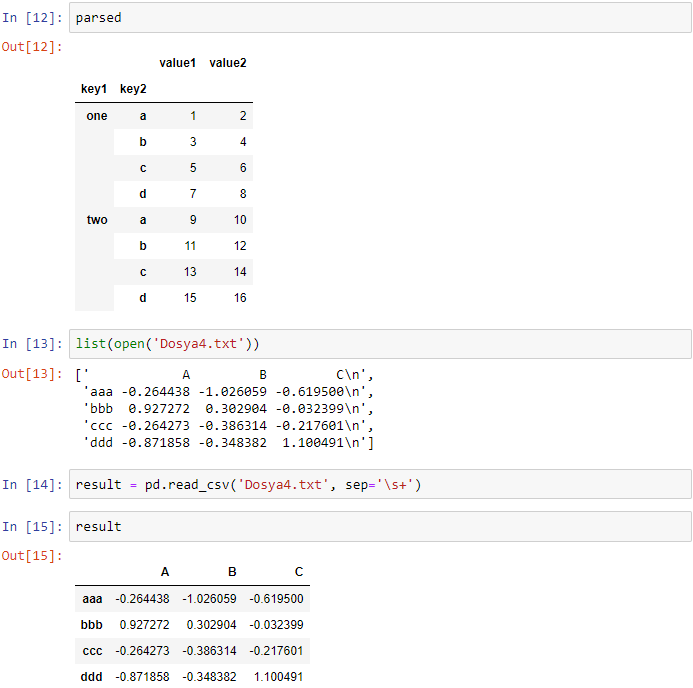
Pandas, tablo verilerini bir DataFrame nesnesi olarak okumak için bir dizi fonksiyona sahiptir.

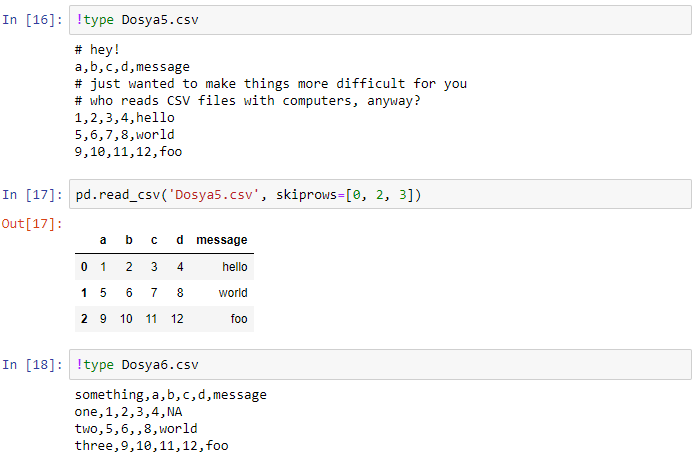
Tablo: Pandas’ta Ayrıştırma Fonksiyonları

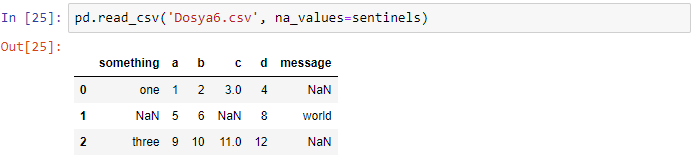
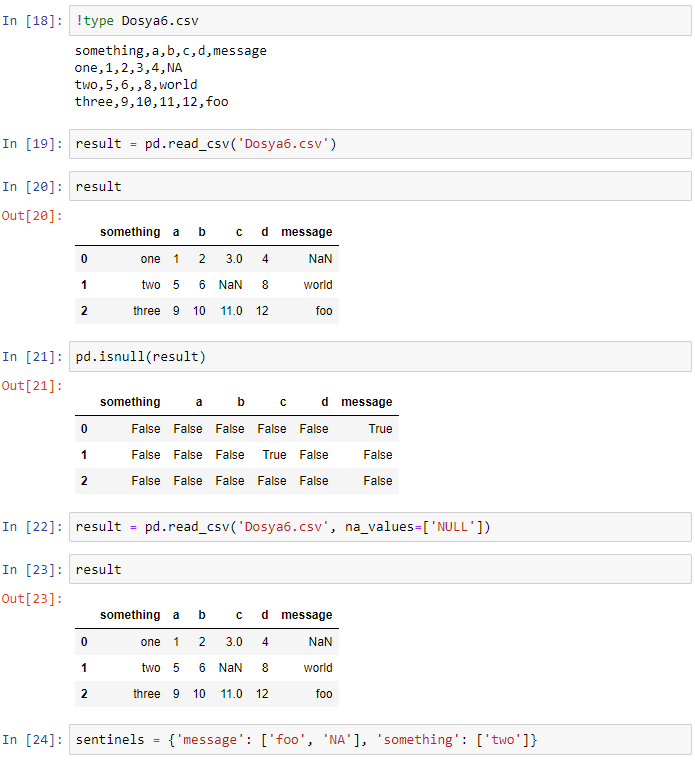
|  |  |
| --- | --- |
| Fonksiyon | Açıklama |
| read\_csv | Ayrılmış veriyi bir dosyadan, URL'den veya dosya benzeri bir nesneden yükler. Varsayılan ayırıcı olarak virgül kullanılır. |
| read\_table | Ayrılmış veriyi bir dosyadan, URL'den veya dosya benzeri bir nesneden yükler. Varsayılan ayırıcı olarak sekme ('\t') kullanılır. |
| read\_fwf | Verileri sabit genişlikte sütun biçiminde okur. |
| read\_clipboard | Panodan veri okuyan read\_table sürümüdür. Web sayfalarından tabloları dönüştürmek için yararlıdır. |
| read\_excel | Excel XLS veya XLSX dosyasındaki tablo verilerini okur. |
| read\_hdf | Pandas tarafından yazılmış HDF5 dosyalarını okur. |
| read\_html | Verilen HTML belgesinde bulunan tüm tabloları okur. |
| read\_json | Bir JSON (JavaScript Object Notation) dizgi gösteriminde verileri okur. |
| read\_msgpack | MessagePack ikili formatı kullanılarak kodlanan pandas verilerini okur. |
| read\_pickle | Python pickle biçiminde saklanan keyfi bir nesneyi okur. |
| read\_sas | SAS sisteminin özel depolama biçimlerinden birinde depolanan bir SAS veri kümesini okur. |
| read\_sql | Bir SQL sorgusunun sonuçlarını pandas dataframe olarak okur. |
| read\_stata | Stata dosya biçiminde bir veri kümesi okur. |
| read\_feather | Feather ikili dosya formatını okur. |



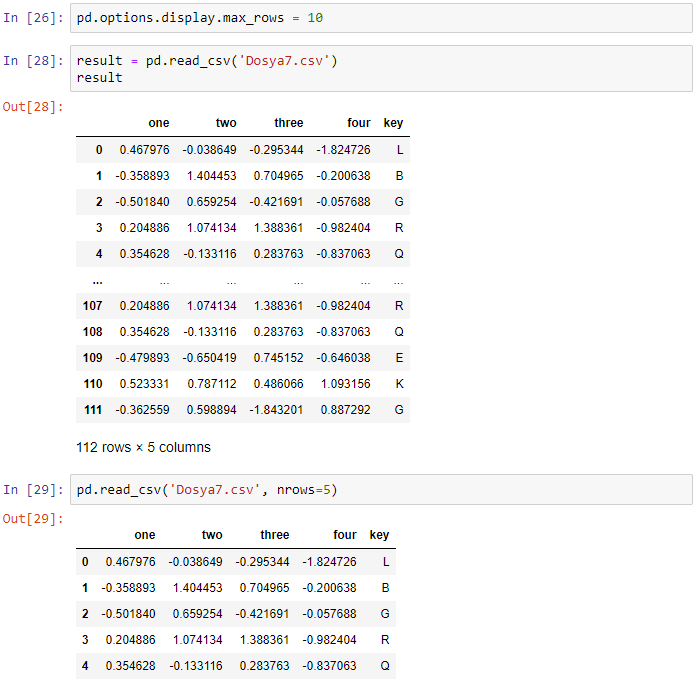




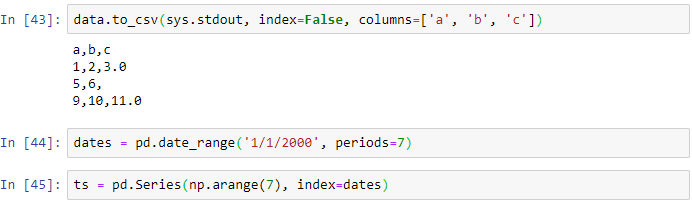
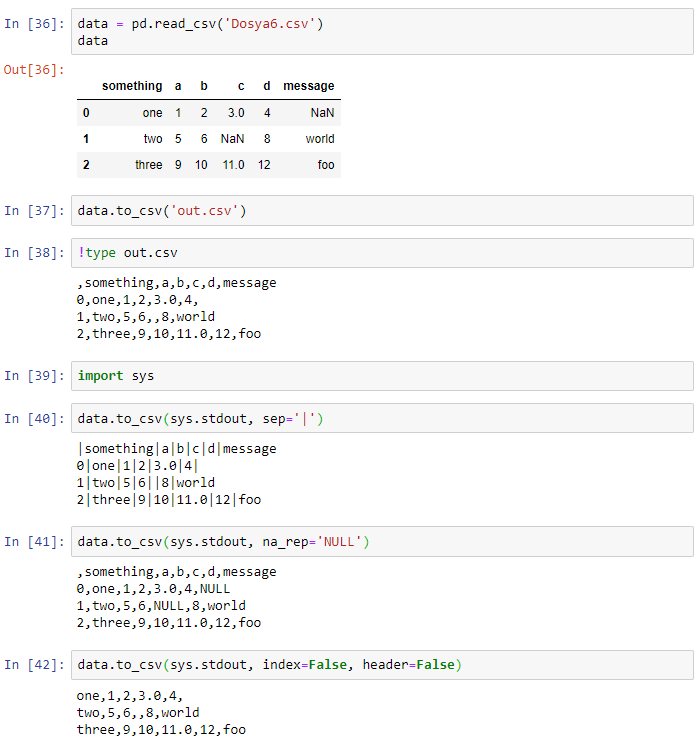




**Metin Dosyalarını Parçalar Halinde Okuma**



**Metin Formatına Veri Yazma**



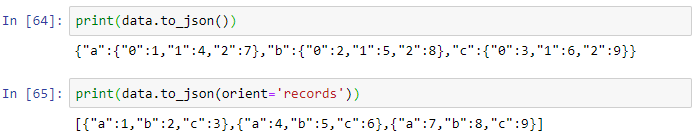
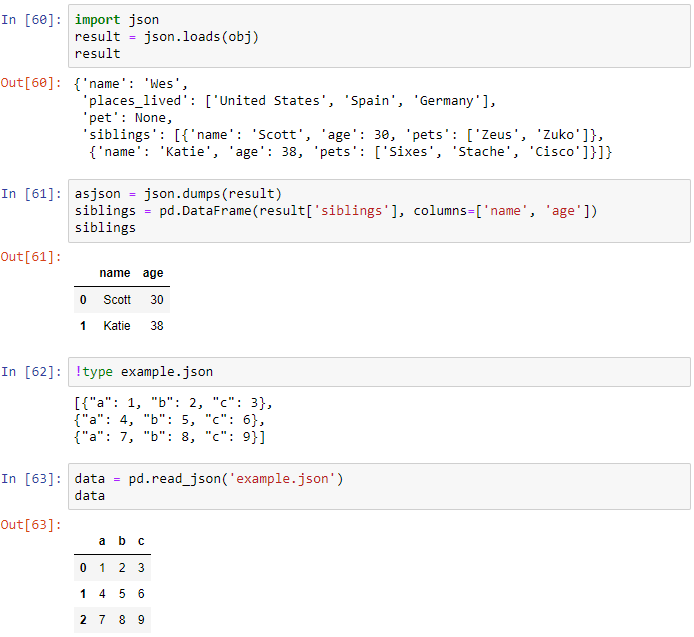


**Sınırlandırılmış Formatlarla Çalışma**



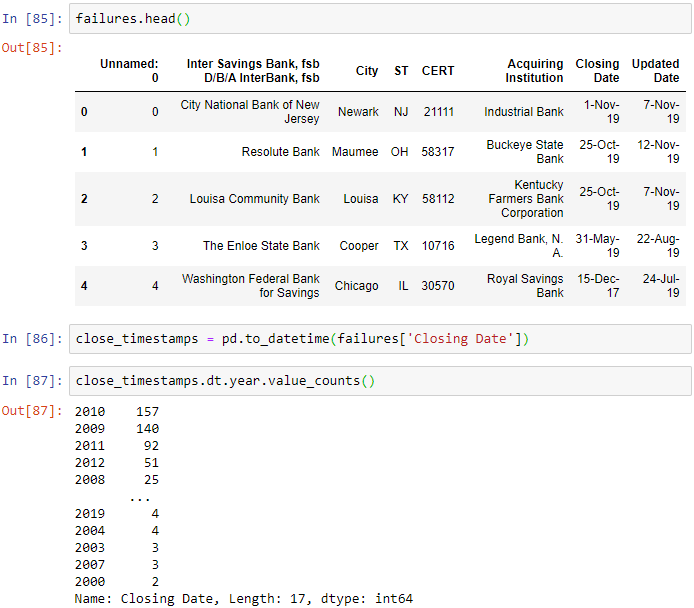
**JSON Verileri**





**XML ve HTML: Web Scraping**





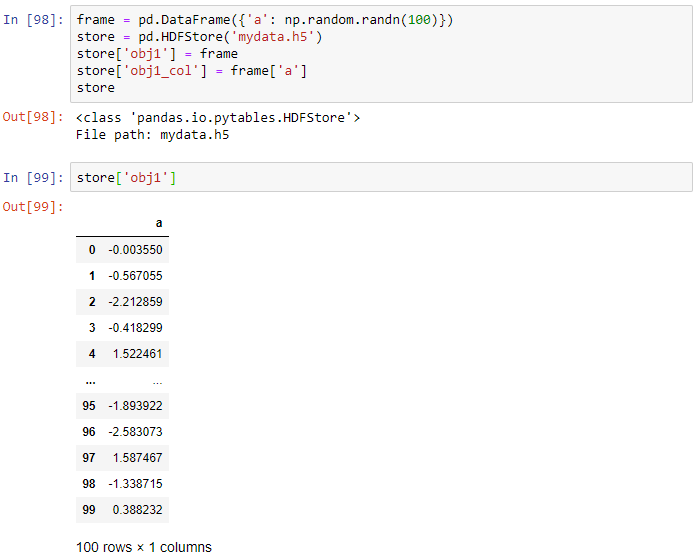
**Lxml.objectify ile XML Ayrıştırma**



**6.2 İkili Veri Formatları**



**HDF5 Formatını Kullanma**





**Microsoft Excel Dosyalarını Okuma**

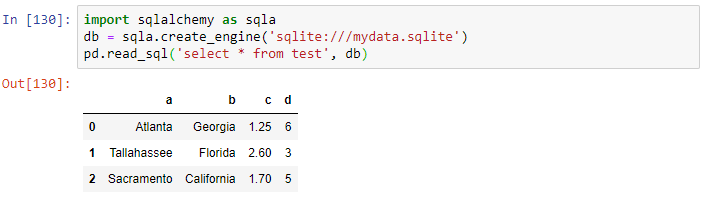


**6.3 Web API'leri ile Etkileşim**



**6.4 Veri Tabanlarıyla Etkileşim**





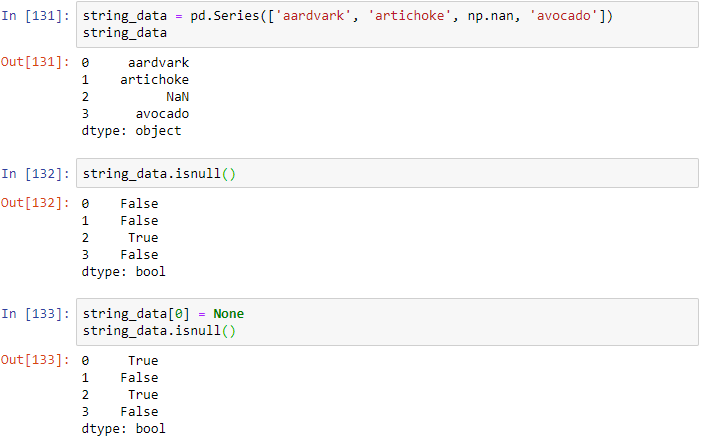
**BÖLÜM 7**

**Veri Temizleme ve Hazırlama**

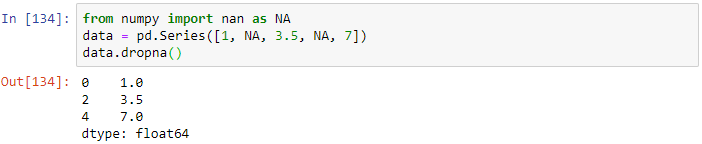
Veri analizi ve modellemesi sırasında veri hazırlığı için önemli miktarda zaman harcanır: yükleme, temizleme, dönüştürme ve yeniden düzenleme. Bu tür görevlerin genellikle bir analistin zamanının % 80'ini veya daha fazlasını üstlendiği bildirilir. Bazen verilerin dosyalarda veya veri tabanlarında saklanma şekli, belirli bir görev için doğru biçimde olmayabilir. Birçok araştırmacı Python, Perl, R veya Java gibi genel amaçlı bir programlama dili veya sed veya awk gibi Unix metin işleme araçlarını kullanarak verilerin bir formdan diğerine geçici olarak işlenmesini tercih eder. Pandas, yerleşik Python dil özellikleri ile birlikte, verileri doğru forma sokmanıza olanak tanıyan yüksek düzeyde, esnek ve hızlı bir araç seti sunar.

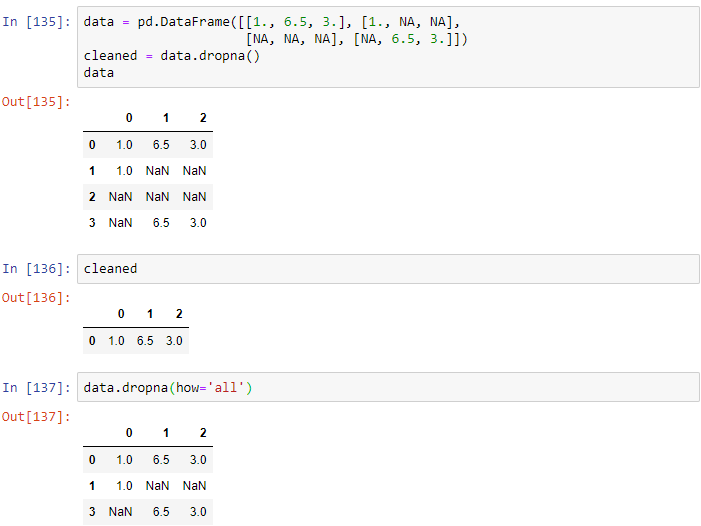
Bu bölümde, eksik veriler, yinelenen veriler, dize manipülasyonu ve diğer bazı analitik veri dönüşümleri için araçlar konu alınmaktadır.

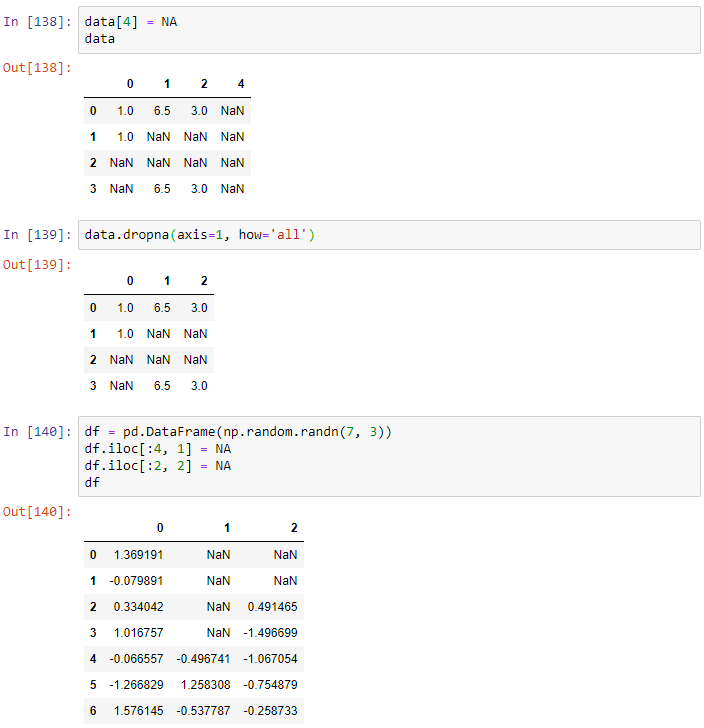
**7.1 Eksik Verileri İşleme**

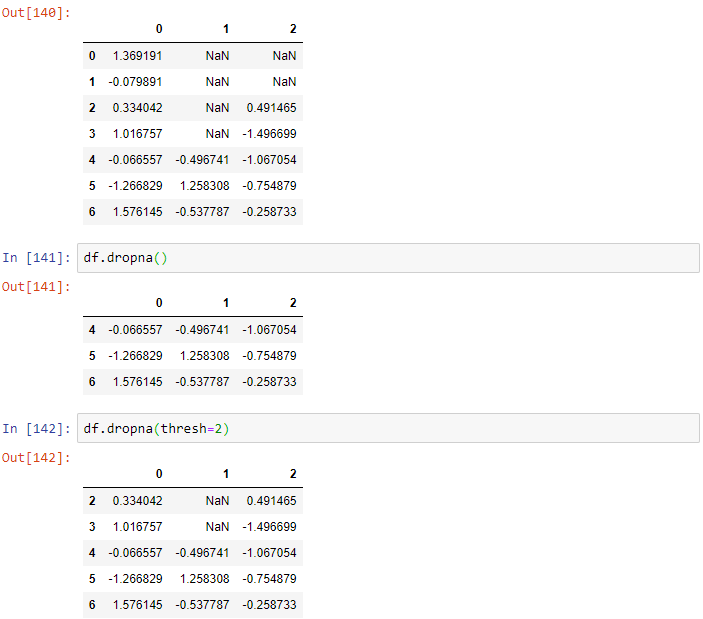


**Eksik Verileri Filtrelemek**

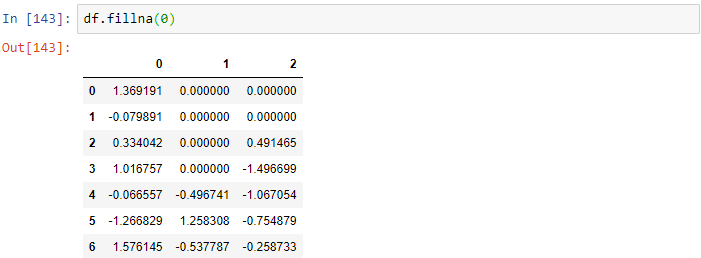


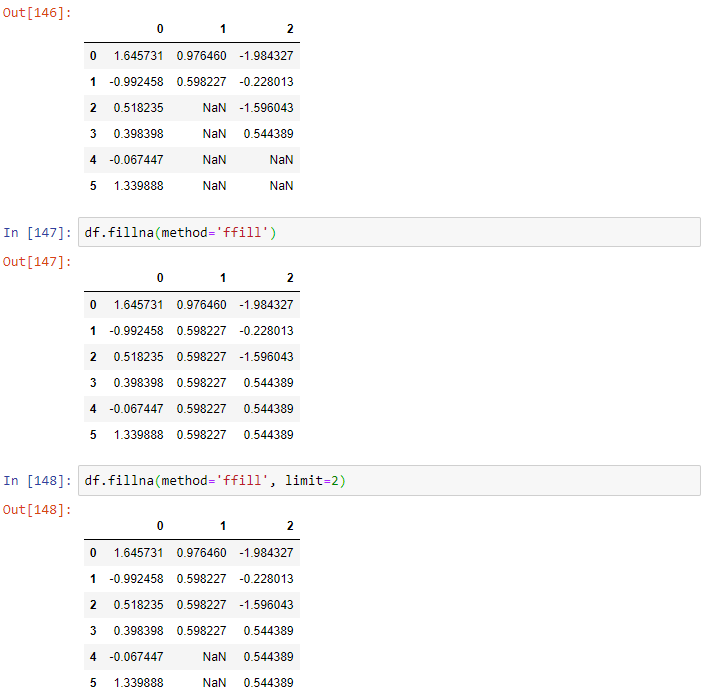
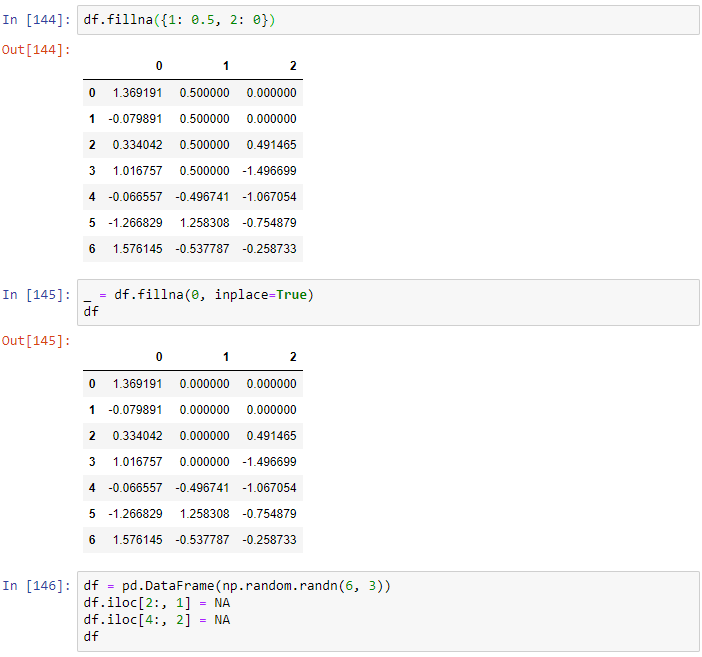


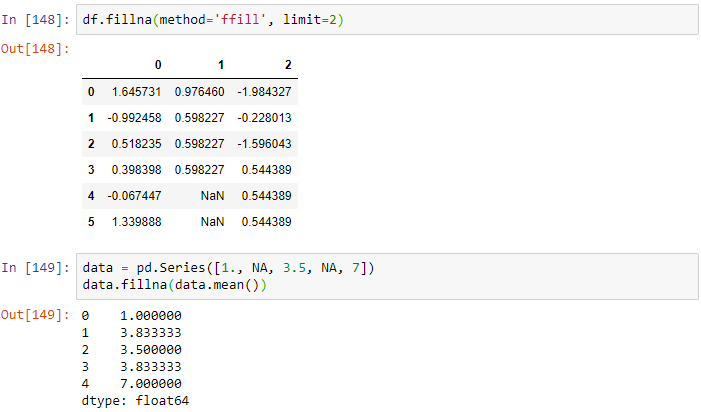




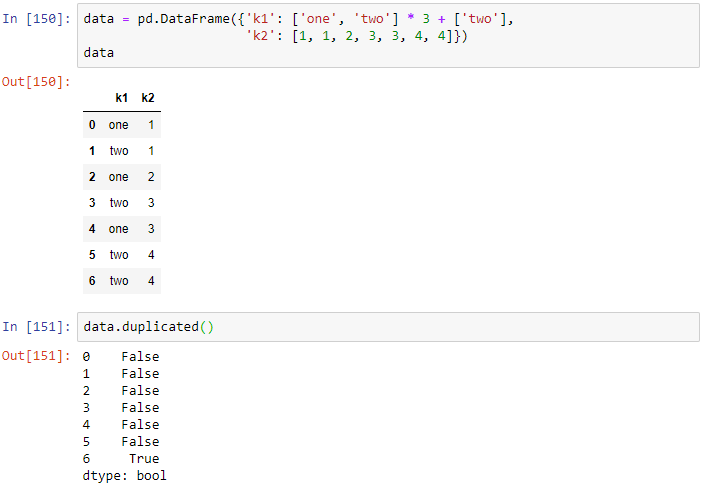
**Eksik Verileri Doldurma**





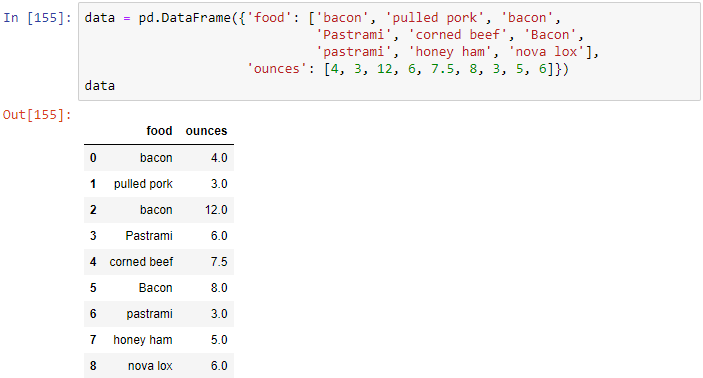


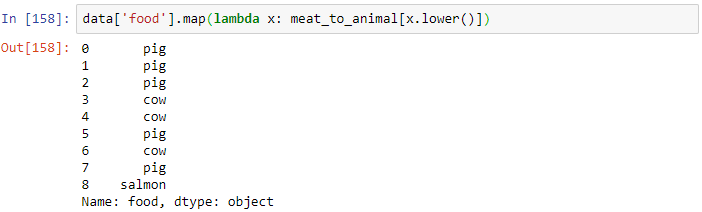
**7.2 Veri Dönüşümü**



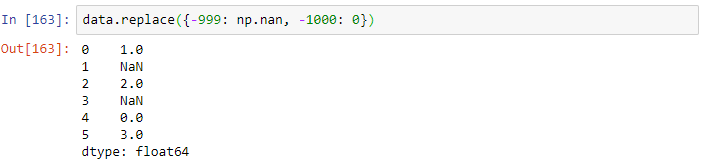
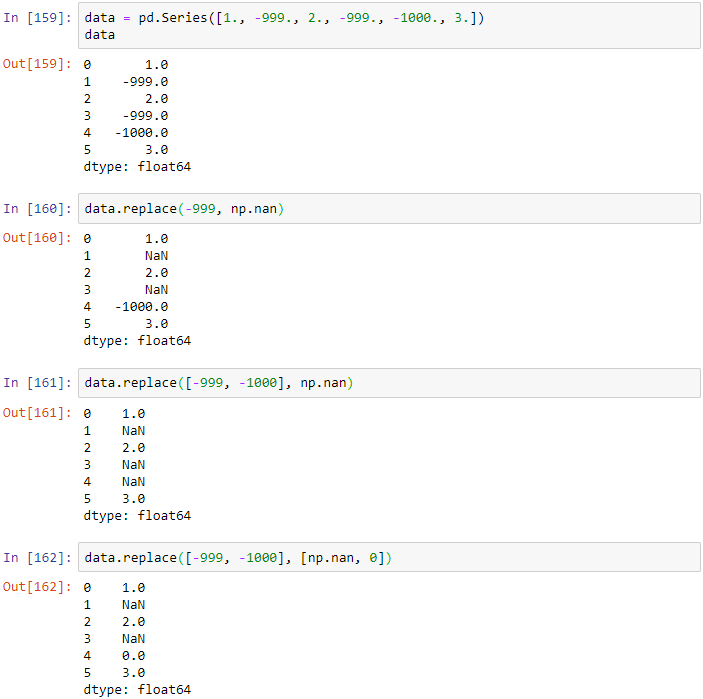


**Bir İşlev veya Eşleme Kullanarak Verileri Dönüştürme**

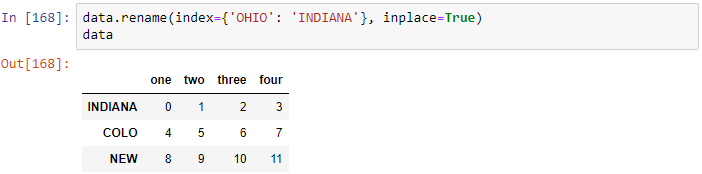




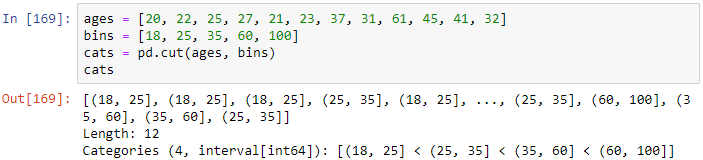
**Değerleri Değiştirme**

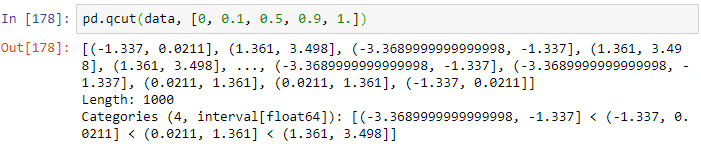
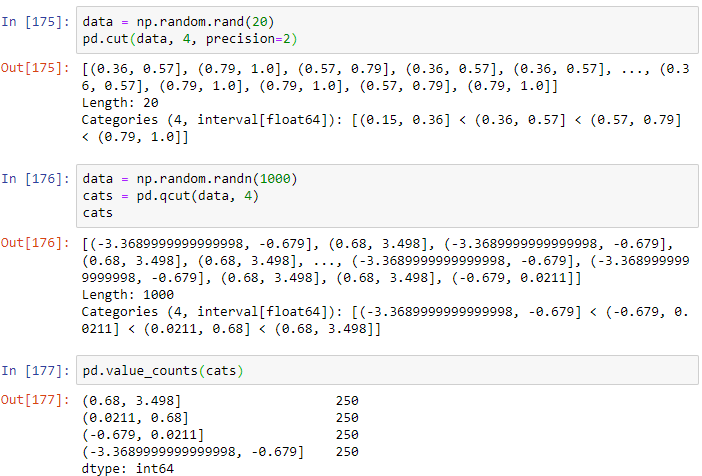
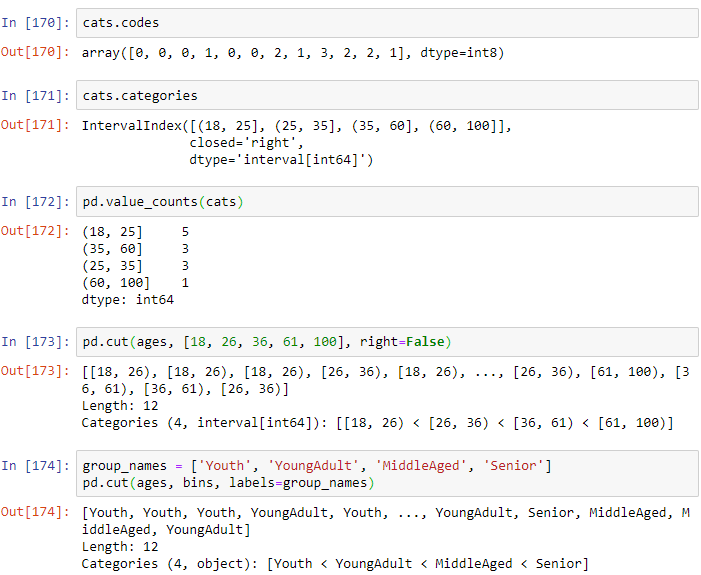


**Eksen Dizinlerini Yeniden Adlandırma**

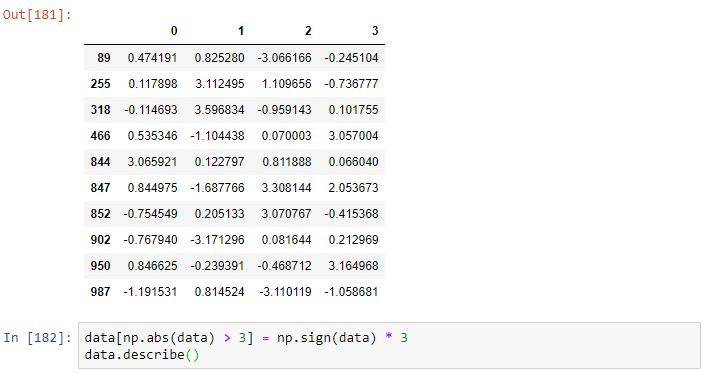
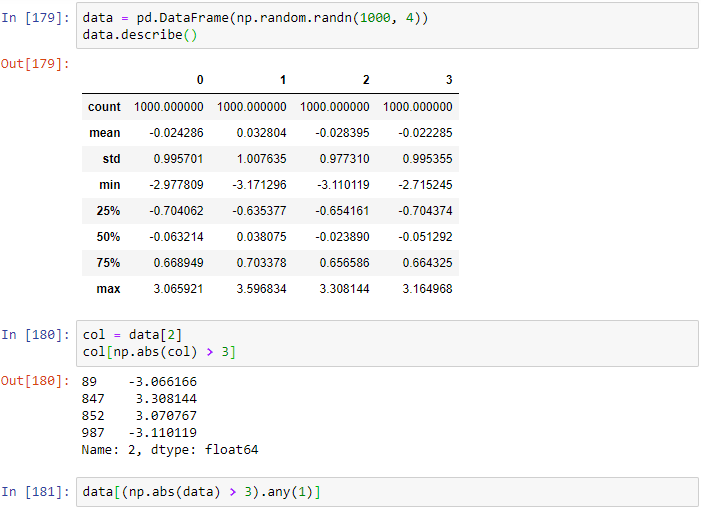


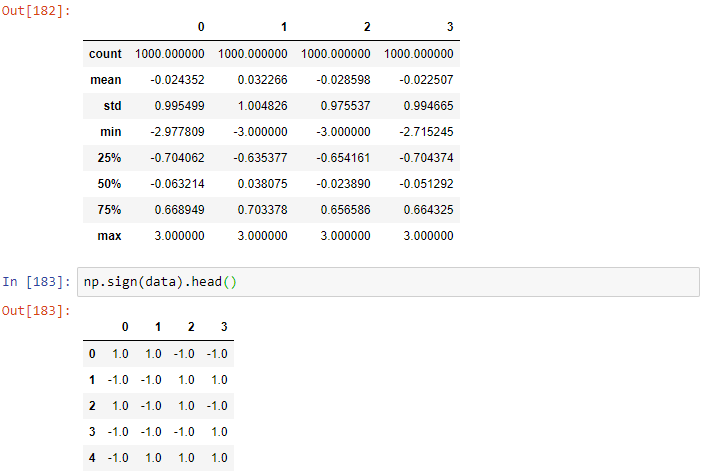
**Discretization ve Binning**



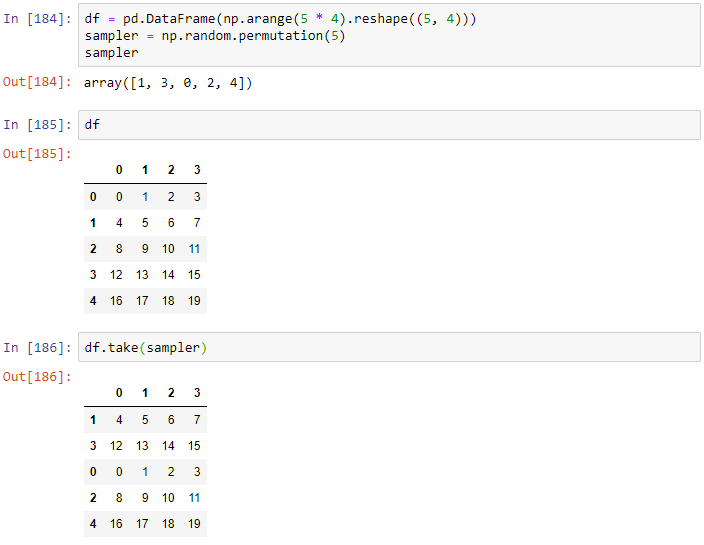


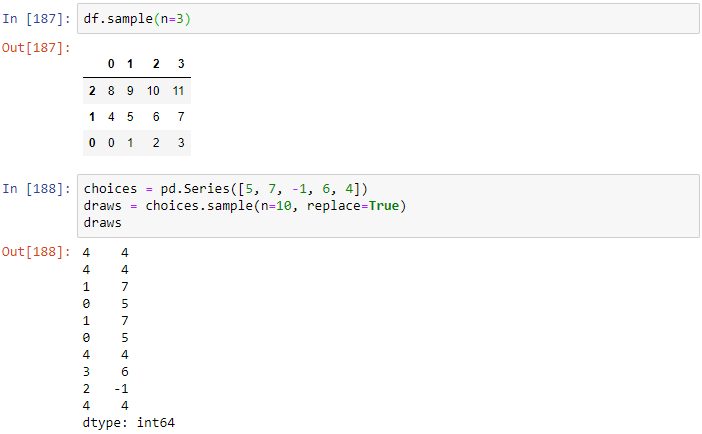
**Aykırı Değerleri Algılama ve Filtreleme**





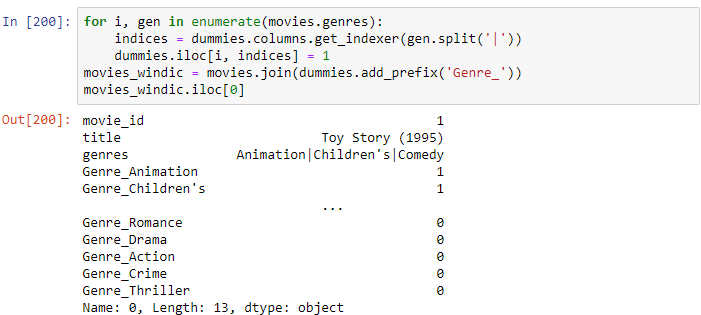
**Permütasyon ve Rastgele Örnekleme**

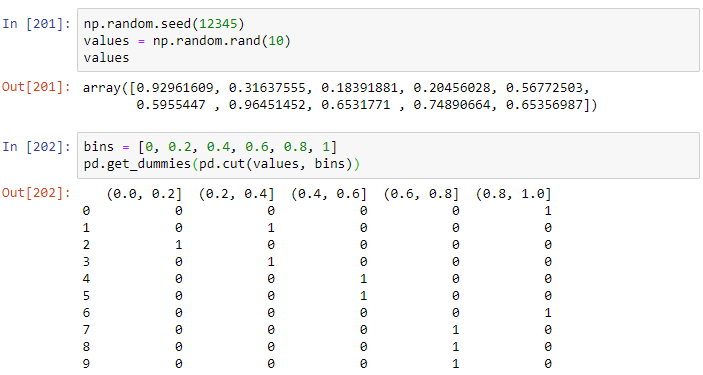


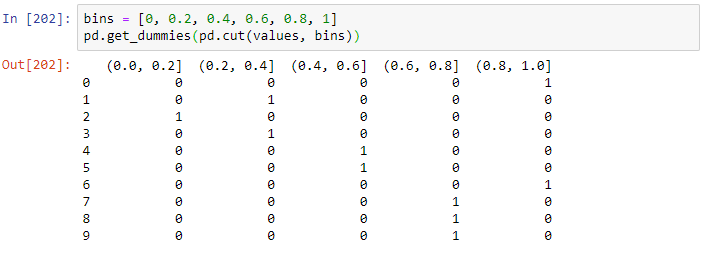


**Hesaplama Göstergesi / Kukla Değişkenler**



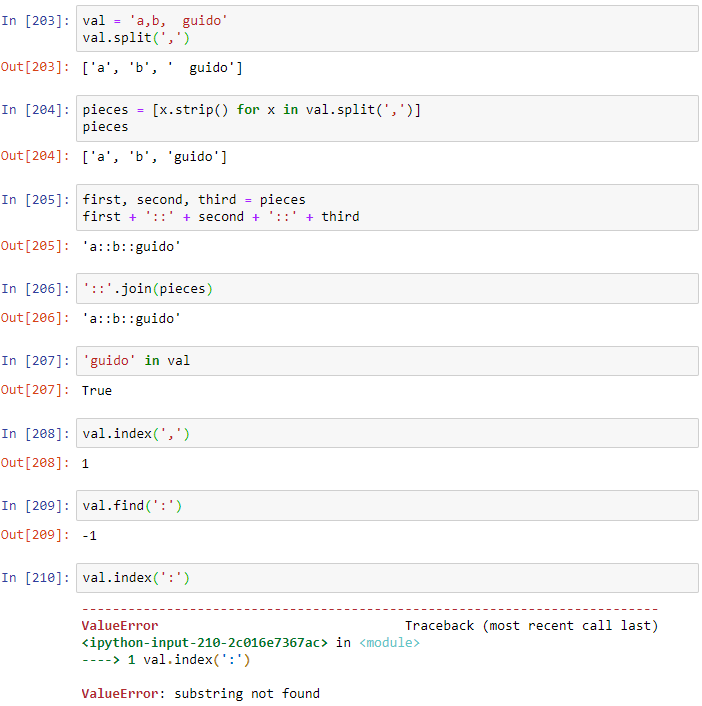






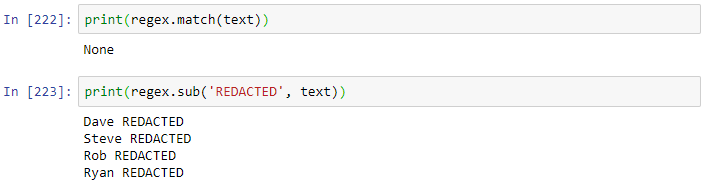
**7.3 String Manipülasyonu**

**String Nesne Yöntemleri**





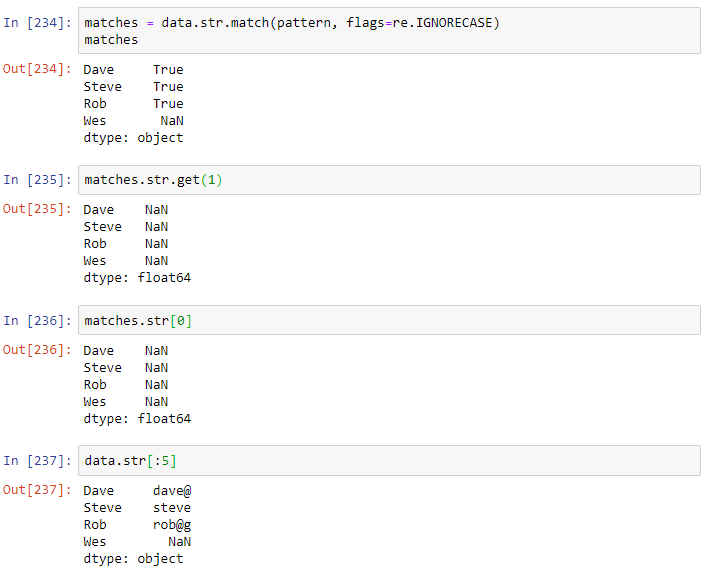
**Düzenli İfadeler (Regular Expressions)**





**Pandas’da Vektörel String İşlemleri**

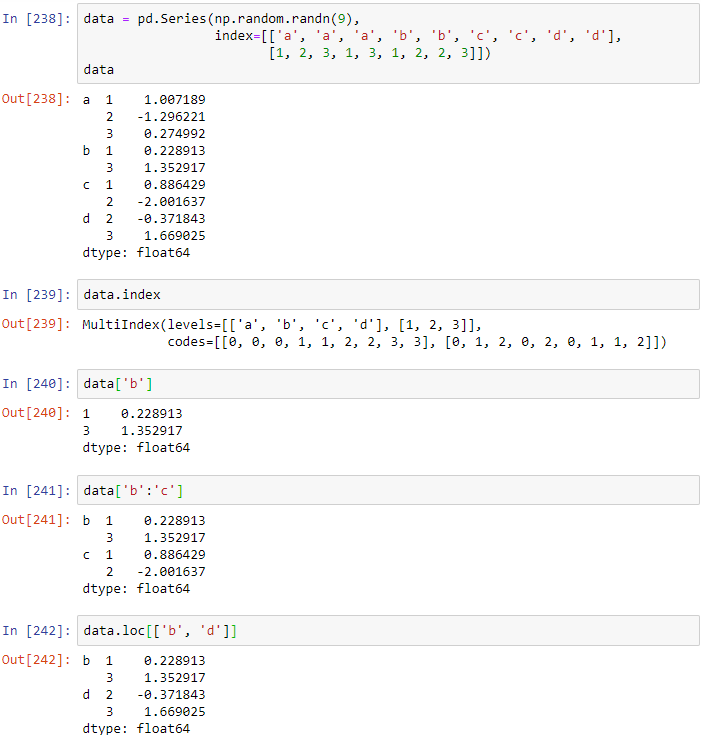


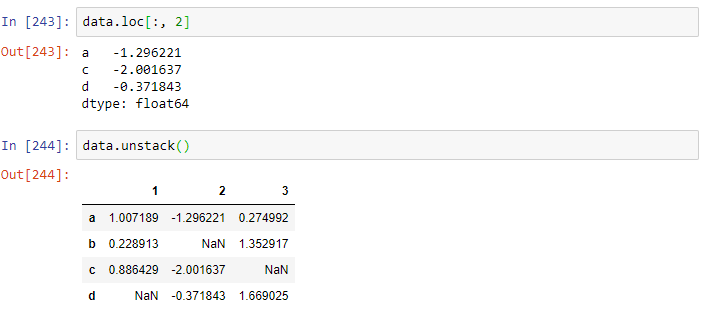


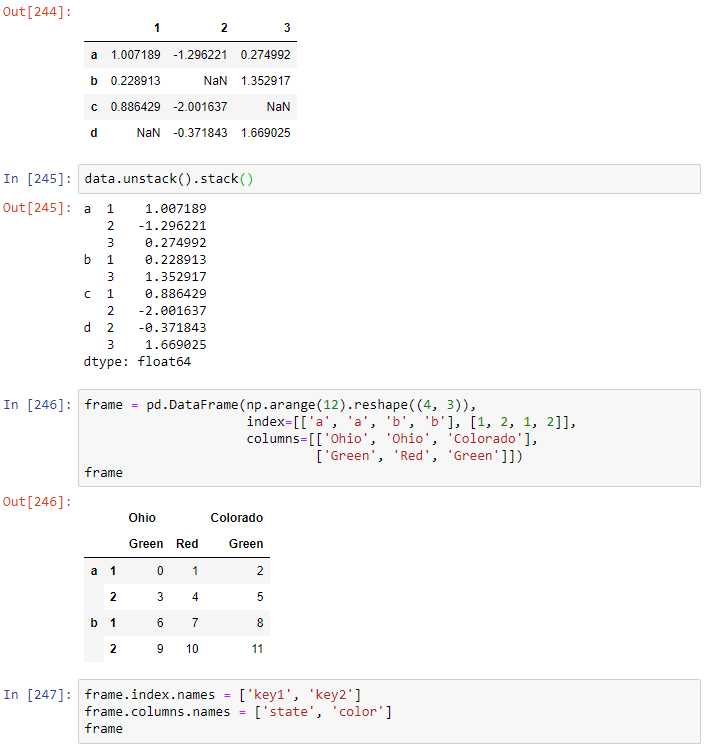
**BÖLÜM 8**

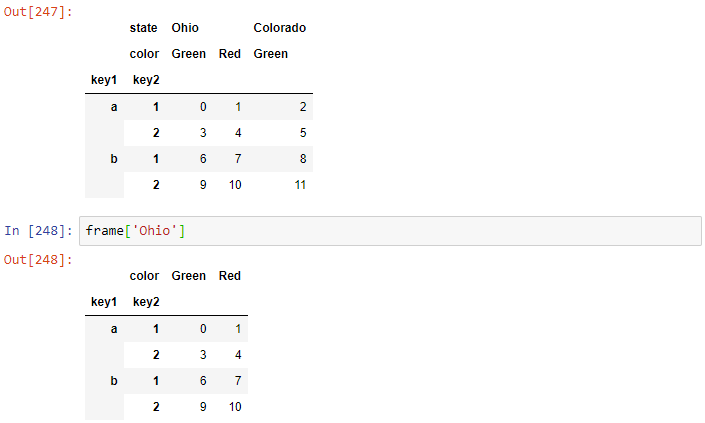
**Veri Kaydırma: Ekle, Birleştir ve Yeniden Şekillendir**

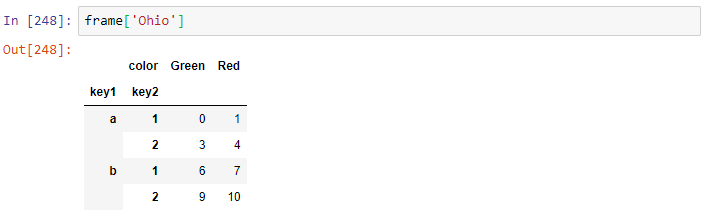
**8.1 Hiyerarşik İndeksleme**



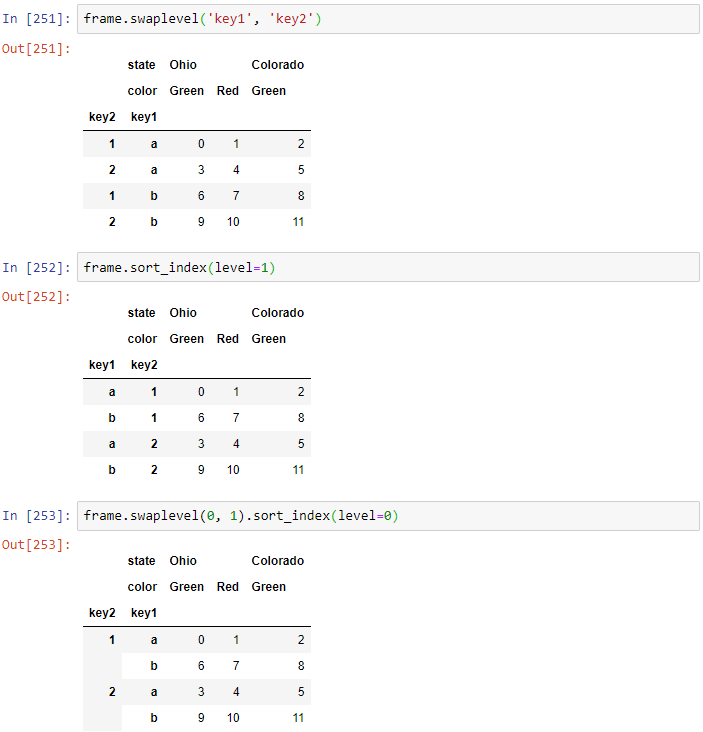




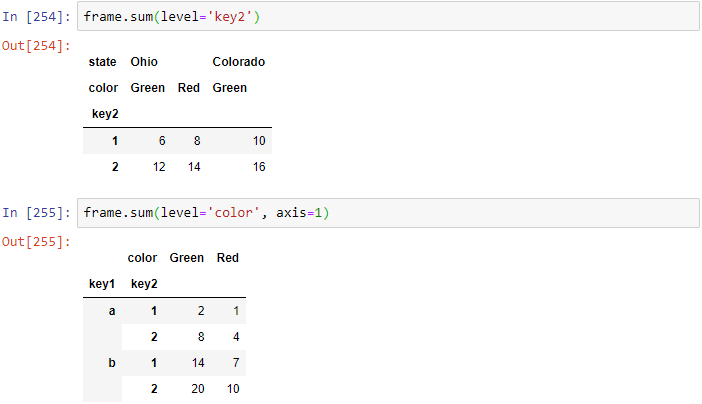




**Yeniden Sıralama ve Sıralama Seviyeleri**

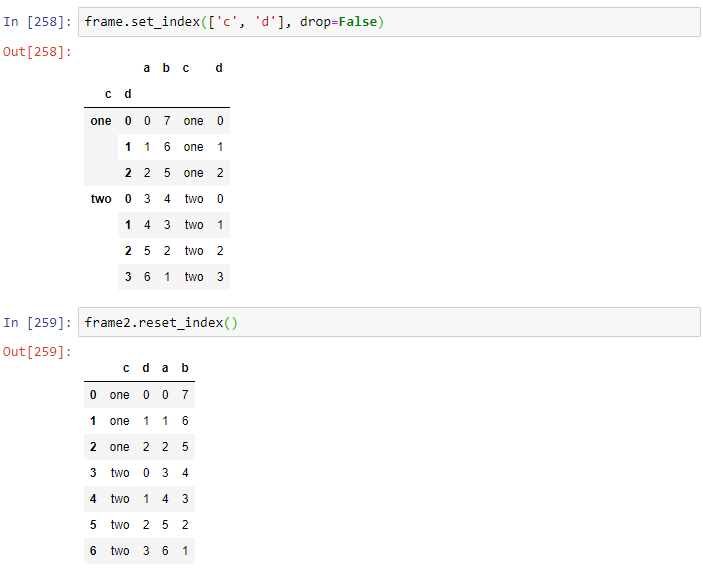


**Seviyeye Göre Özet İstatistikler**



**DataFrame Sütunlarıyla İndeksleme**



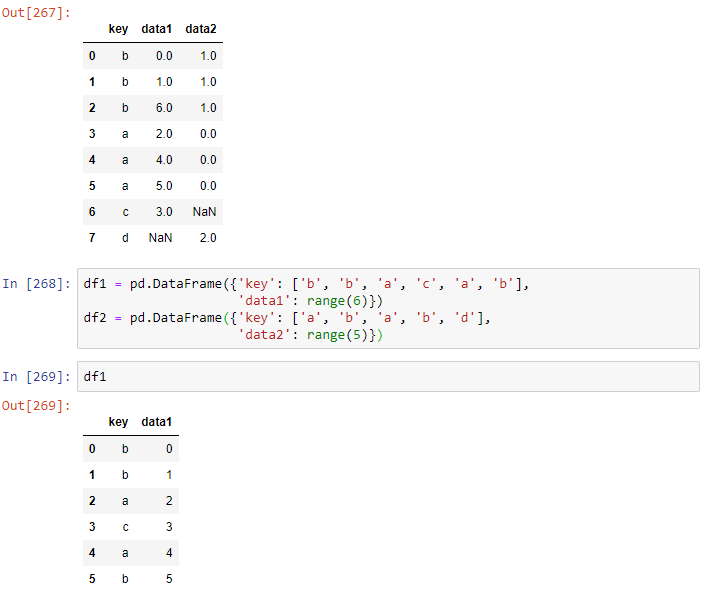


**8.2 Veri Kümelerinin Birleştirilmesi**

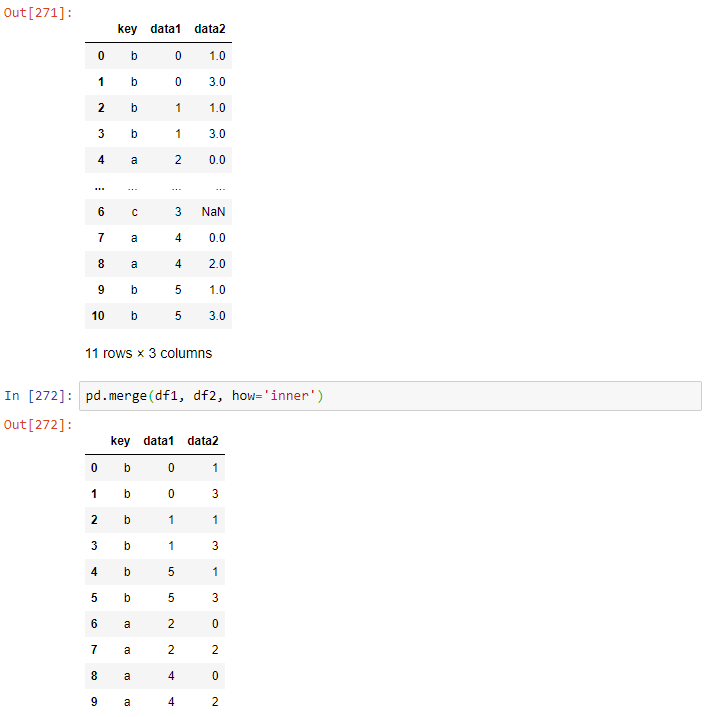


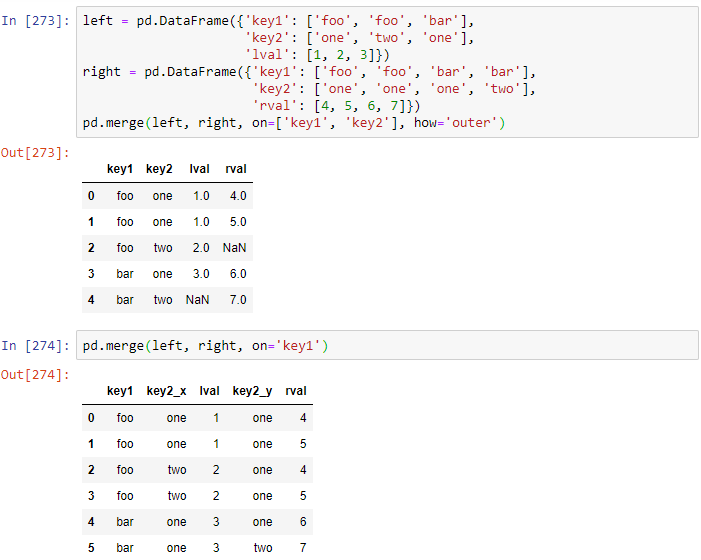


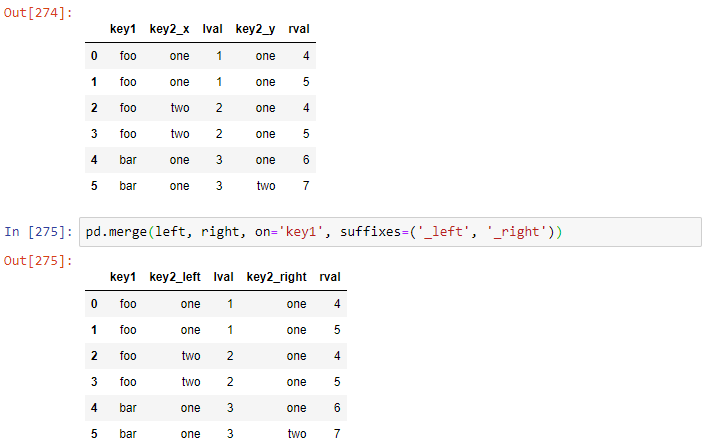






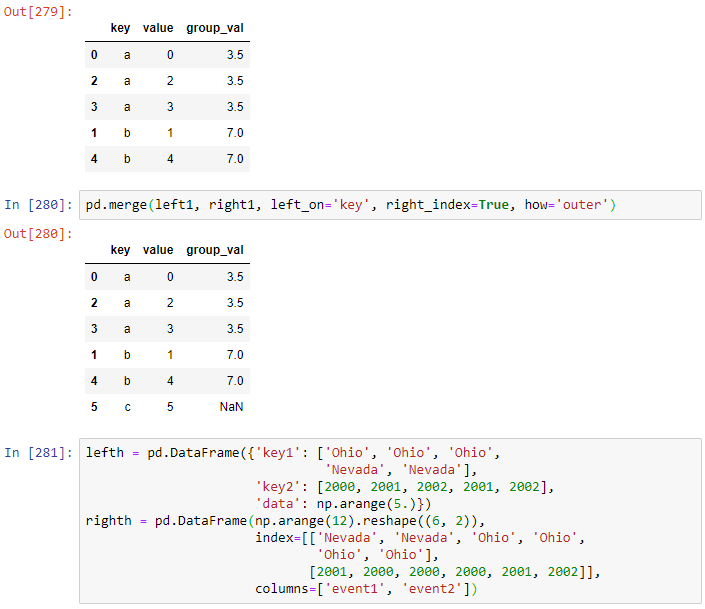






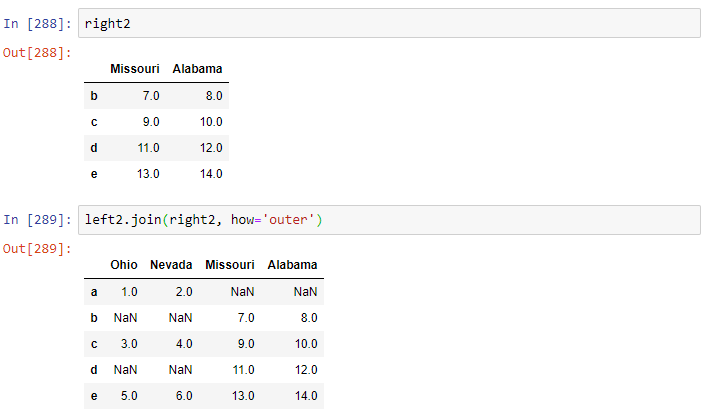
**Dizin Üzerinde Birleştirme**

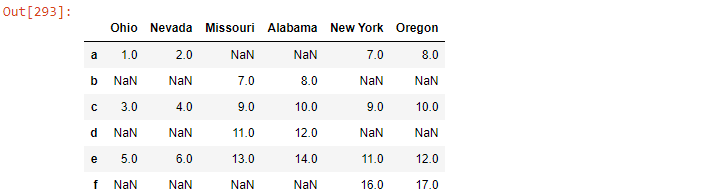
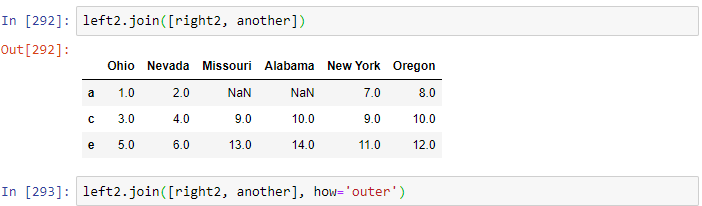
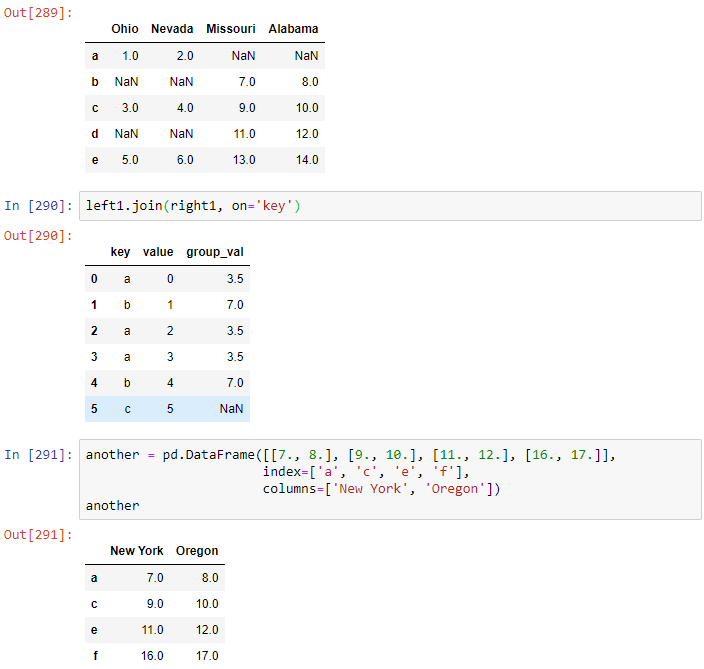




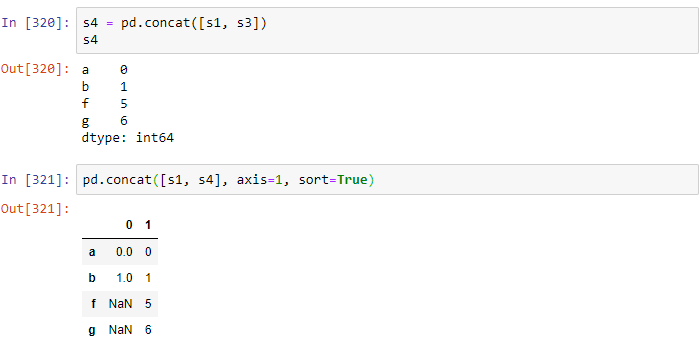
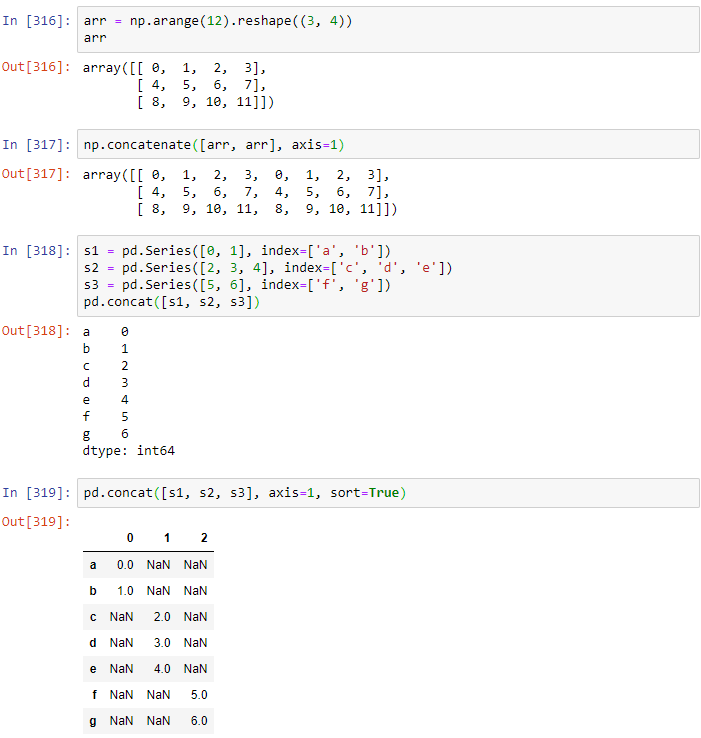


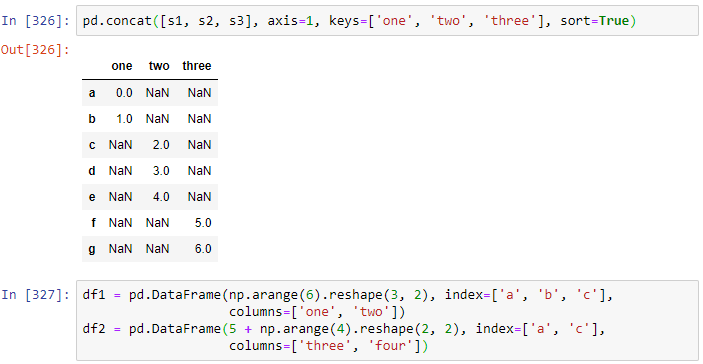
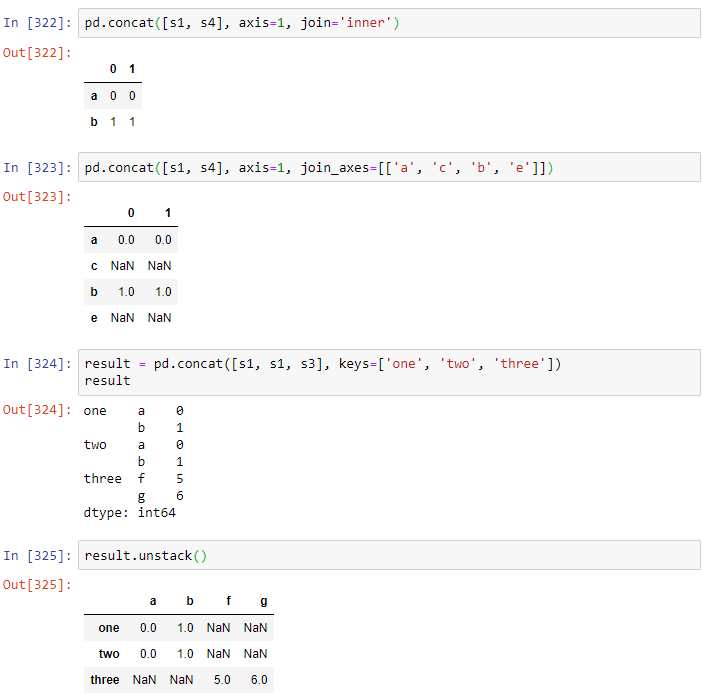


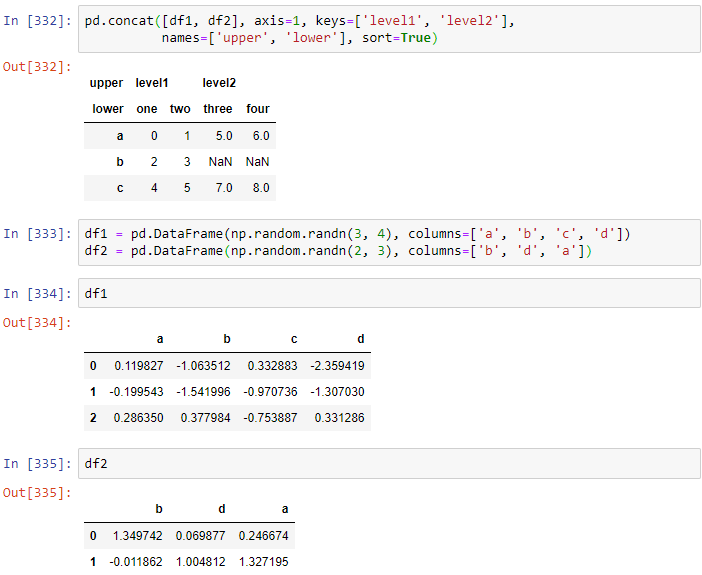
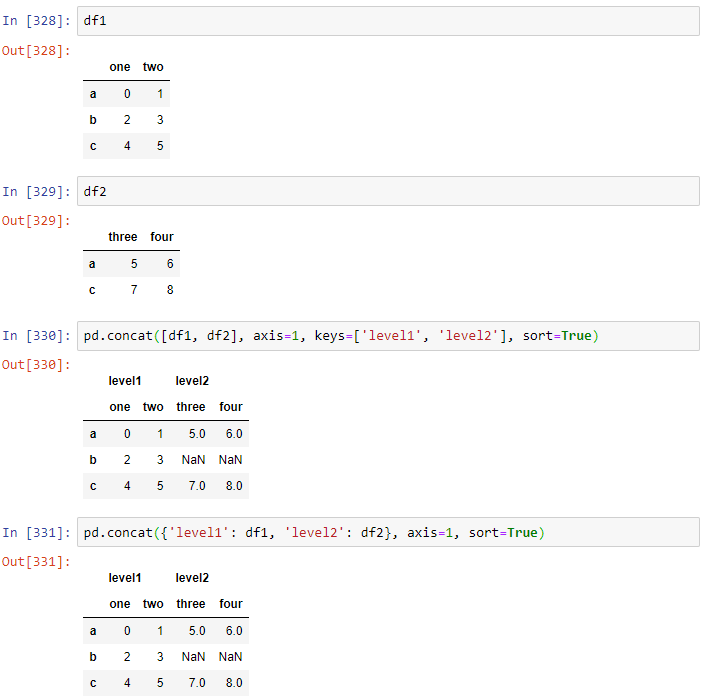


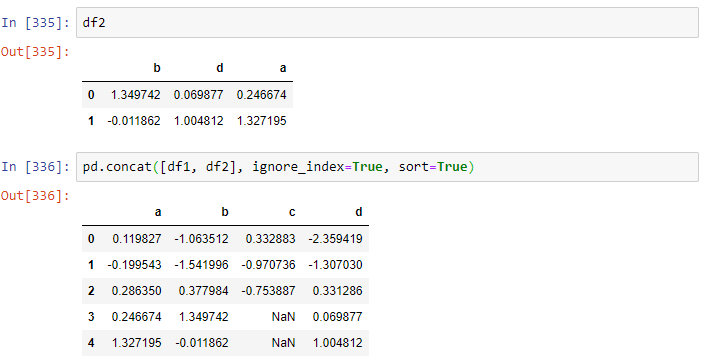


**Bir Eksen Boyunca Birleştirme**

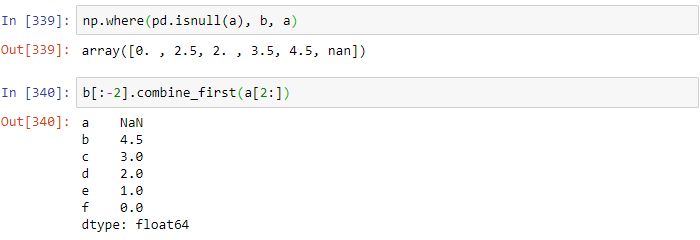
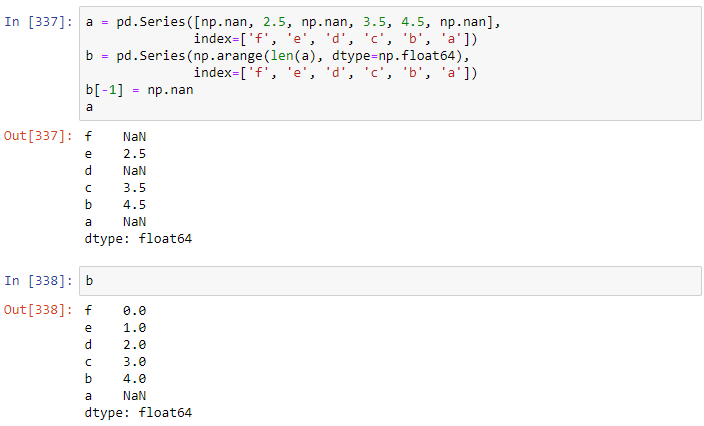


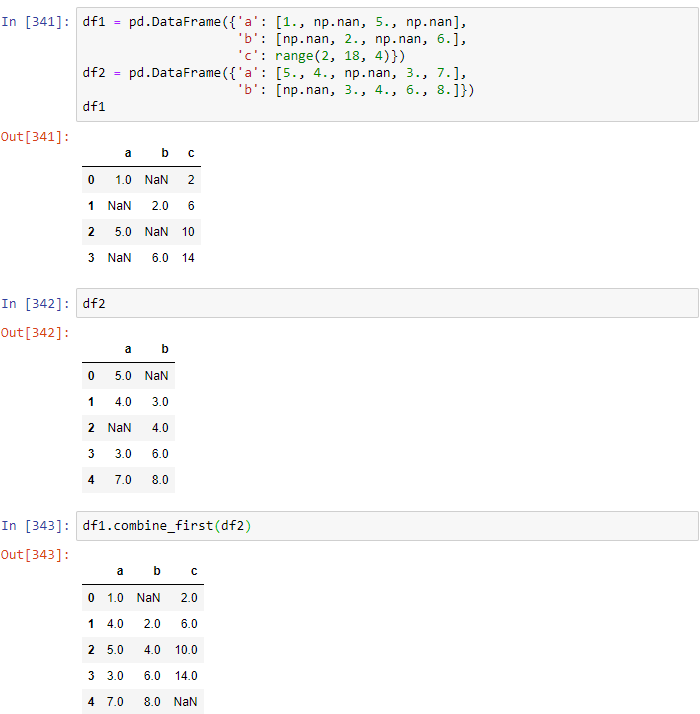






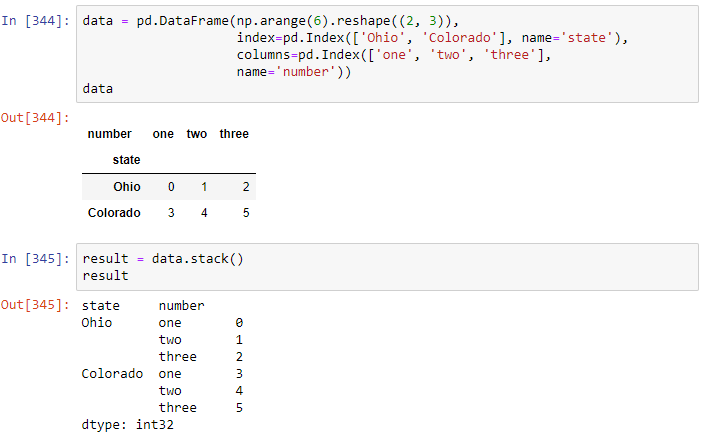
**Örtüşme ile Verileri Birleştirme**

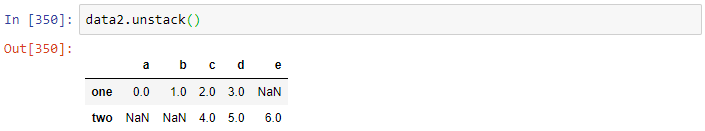
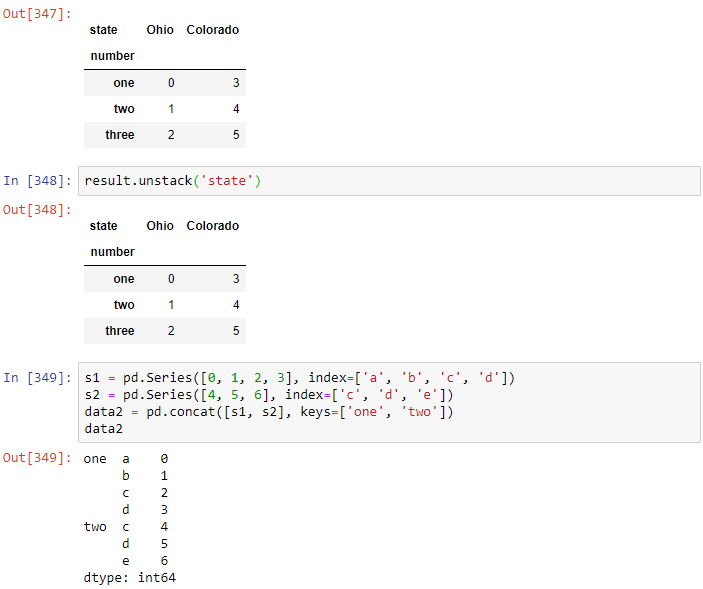
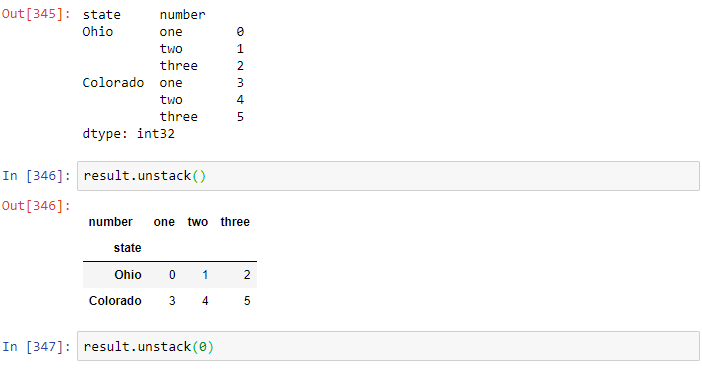


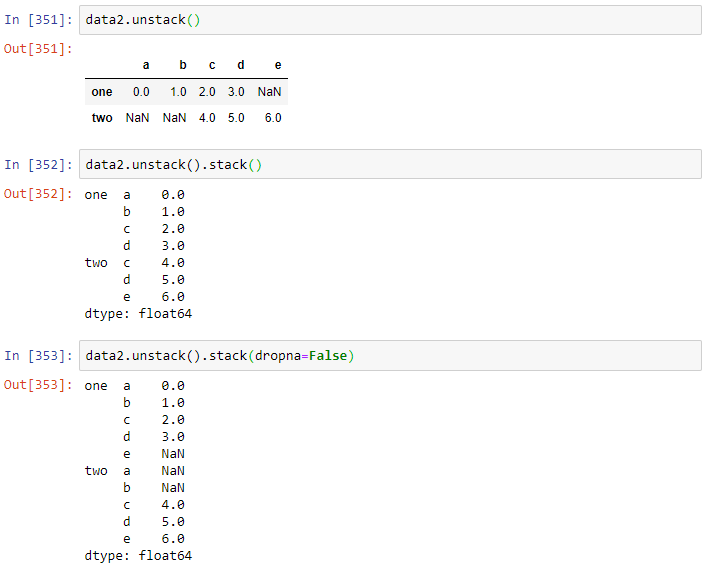


**8.3. Yeniden Şekillendirme ve Döndürme**

**Hiyerarşik İndeksleme ile Yeniden Şekillendirme**

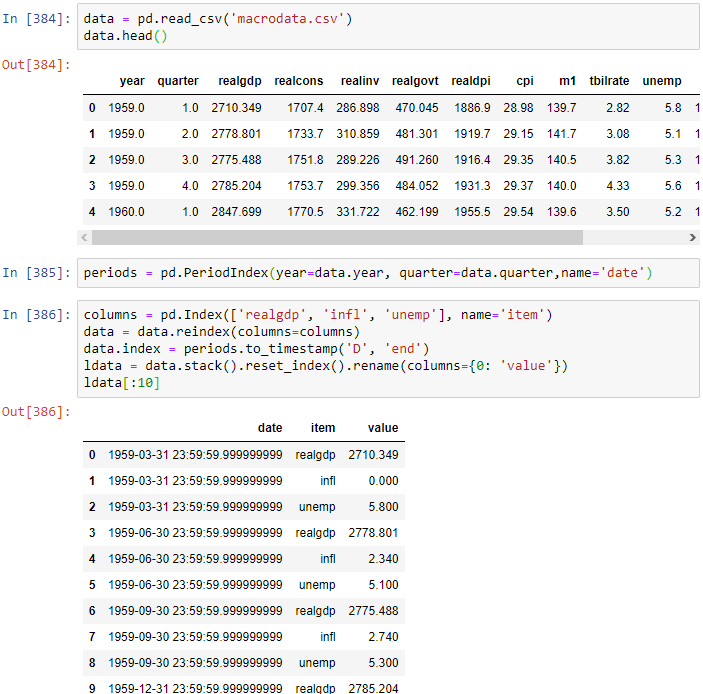


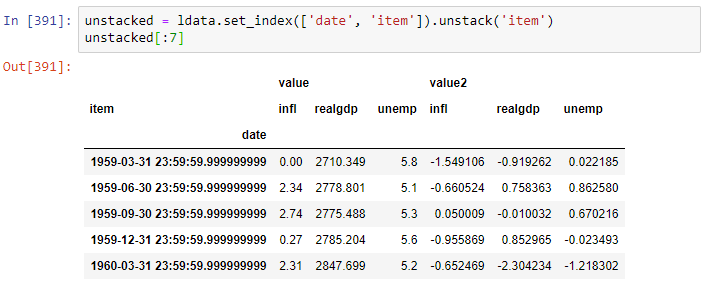
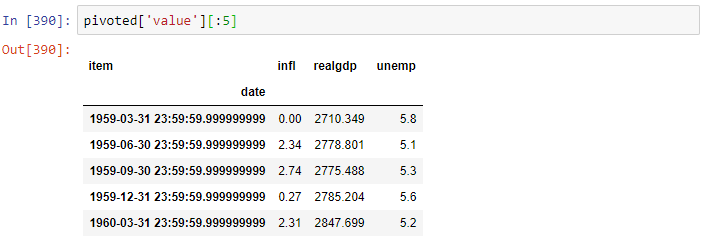
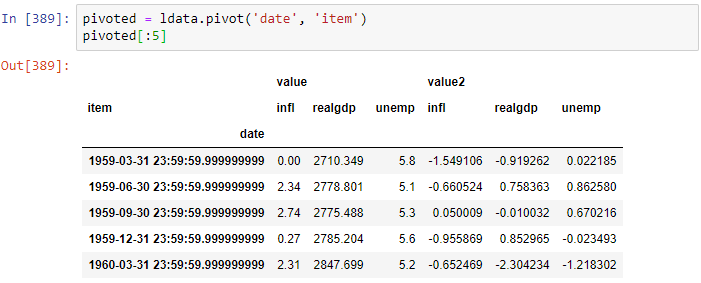






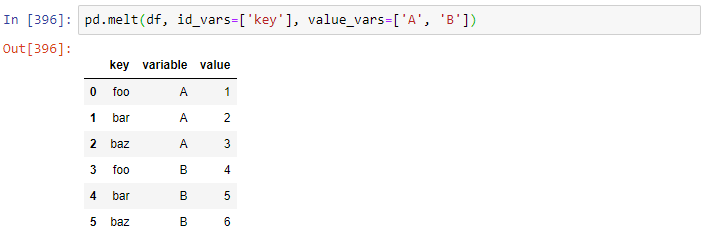
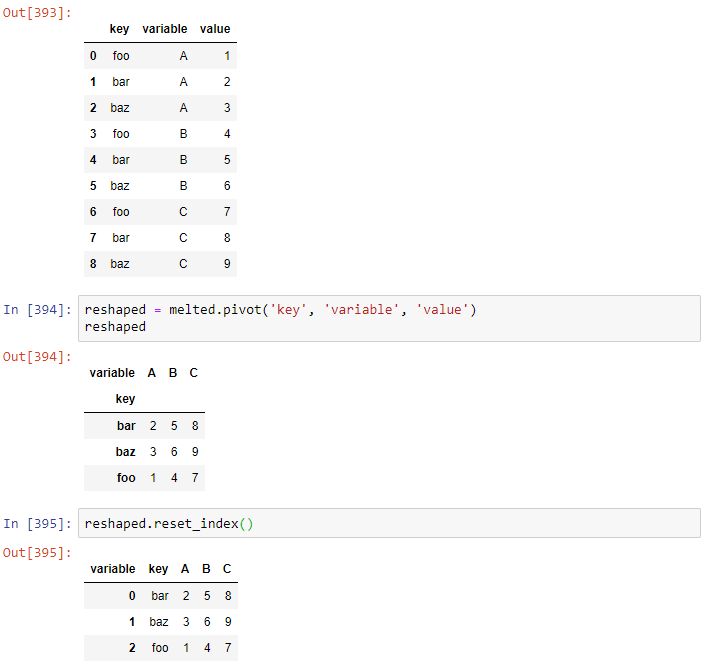
**“Uzun” Biçimden “Geniş” Biçime Dönüştürme**

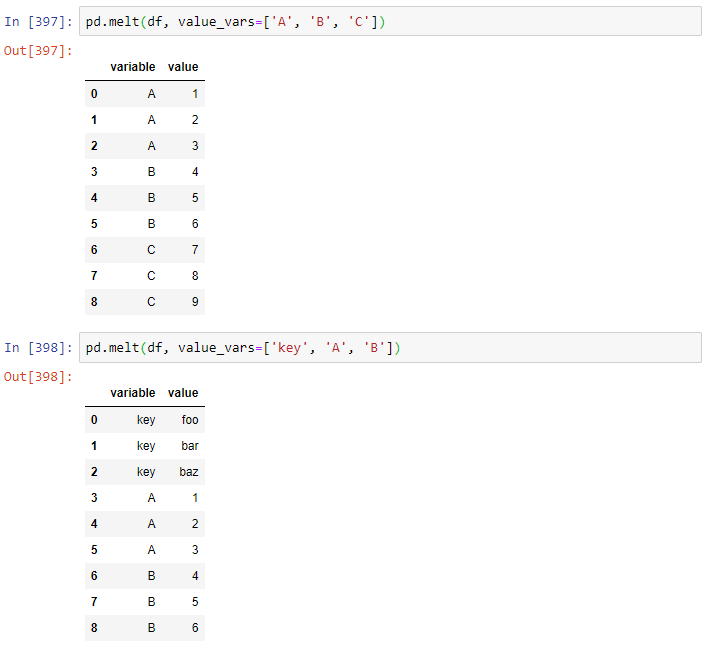




**“Geniş” Biçimden “Uzun” Biçime Dönüştürme**



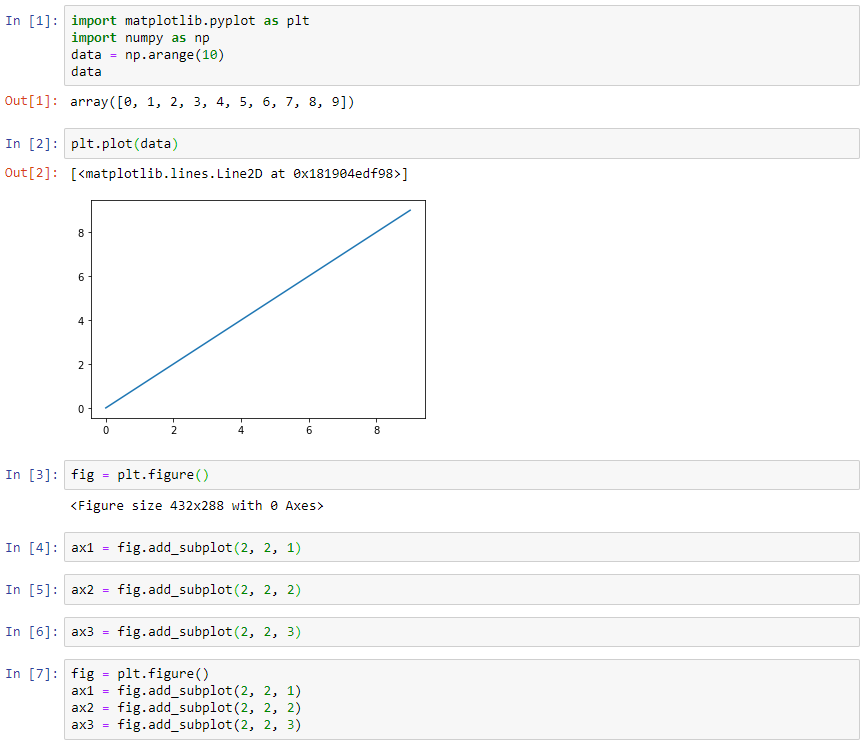


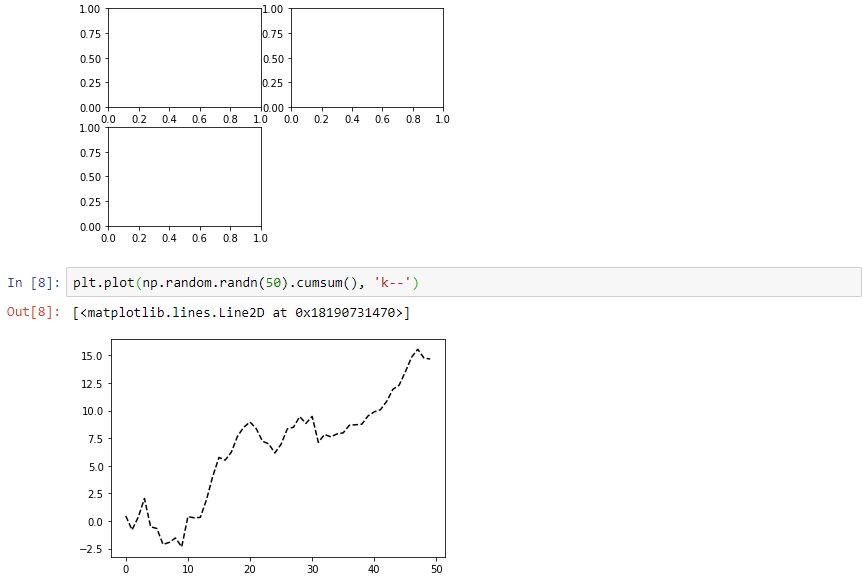


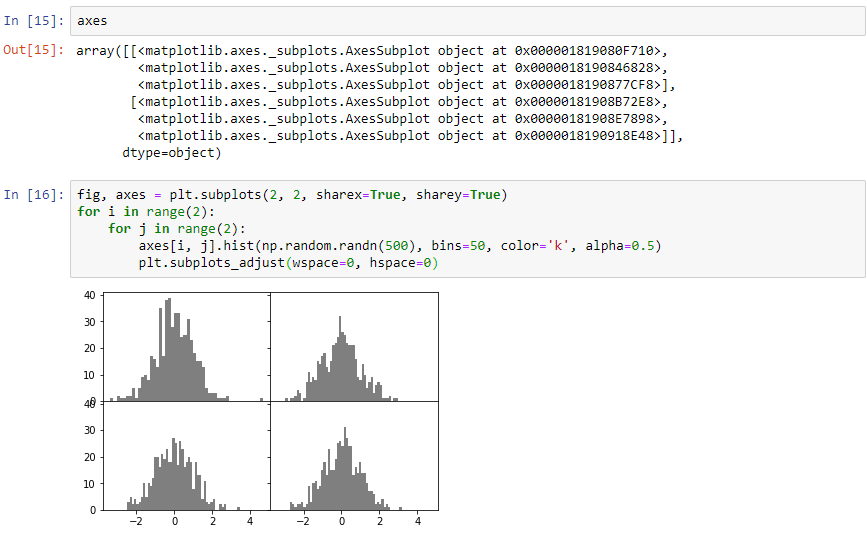
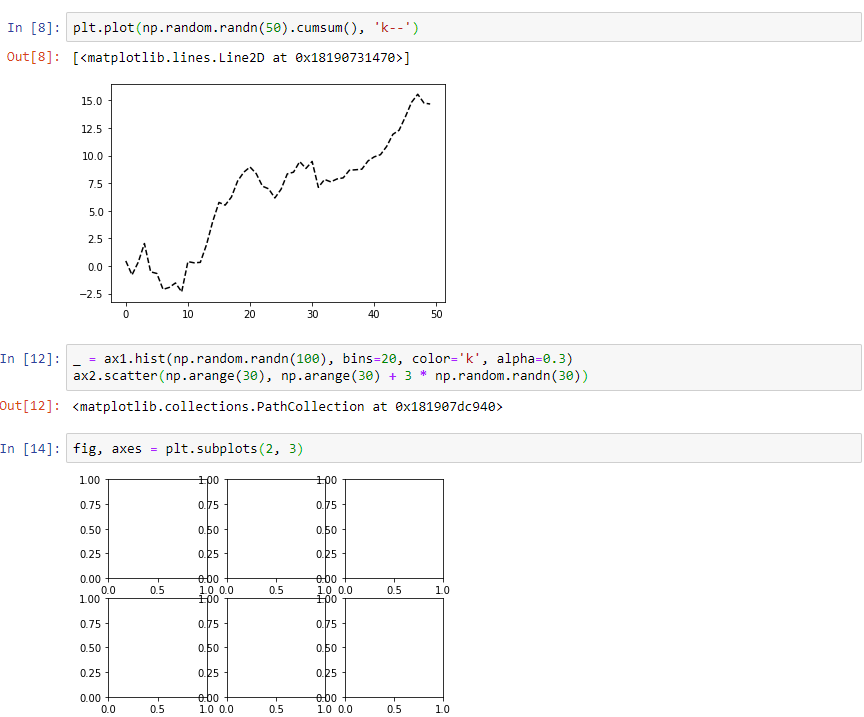
**BÖLÜM 9**

**Çizim ve Görselleştirme**

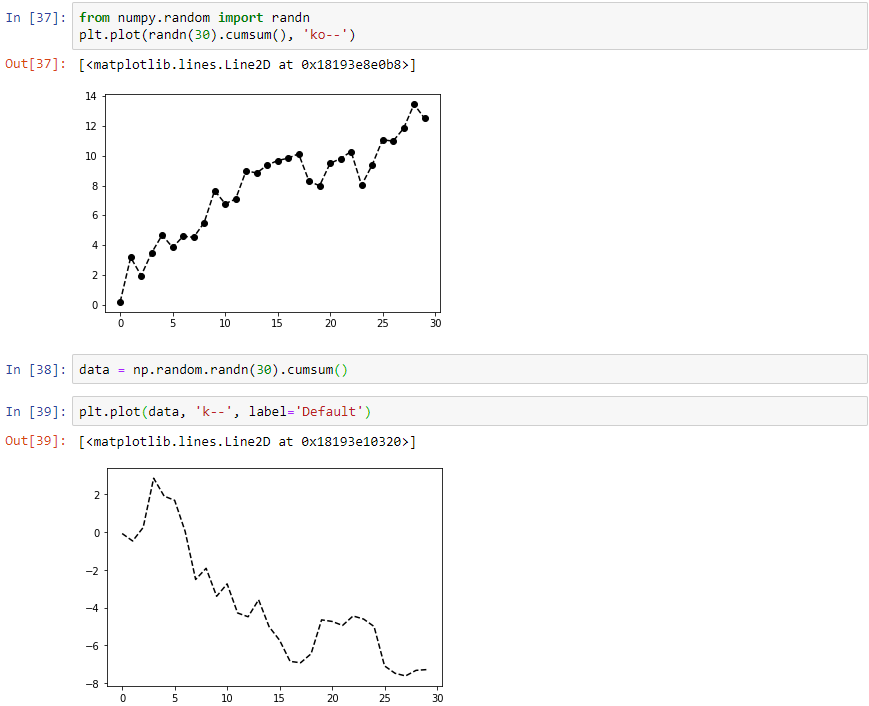
**9.1 Matplotlib**

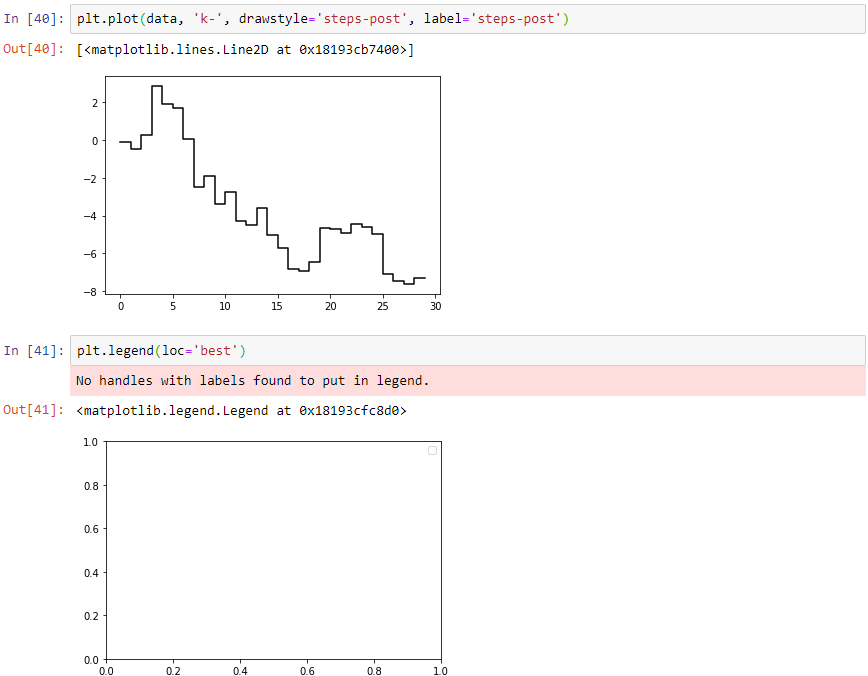


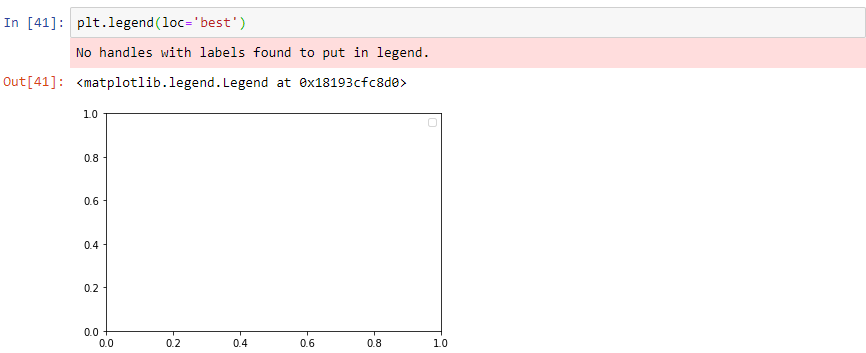




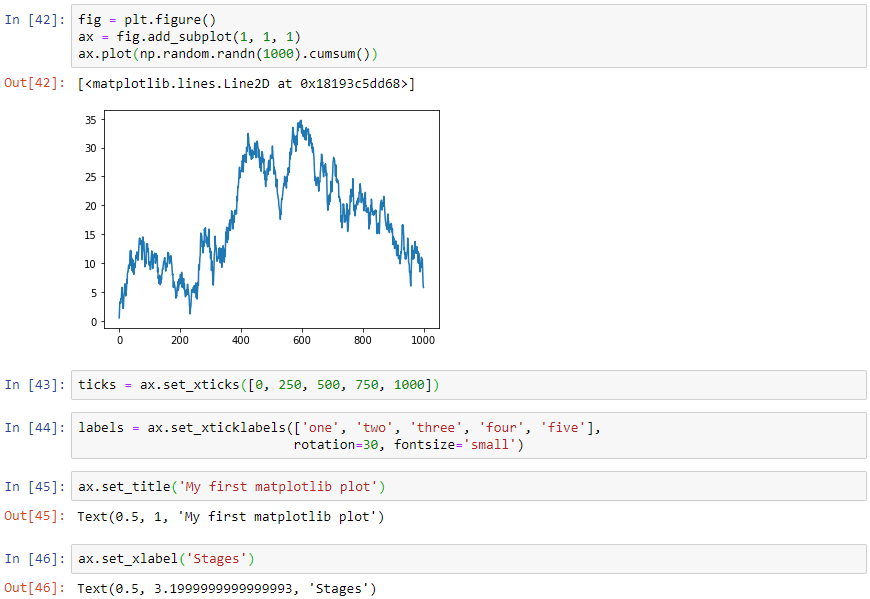
**Renkler, İşaretçiler ve Çizgi Stilleri**





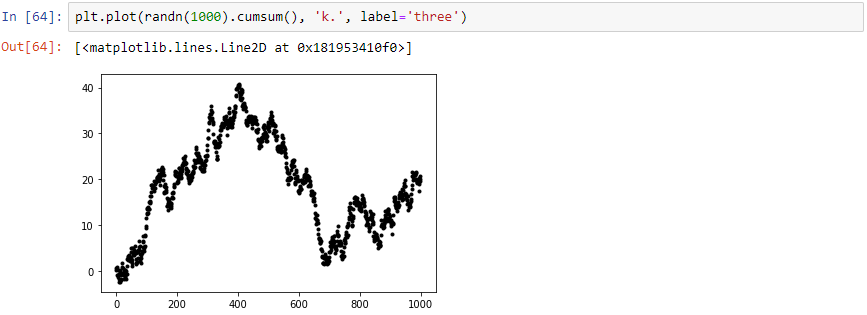
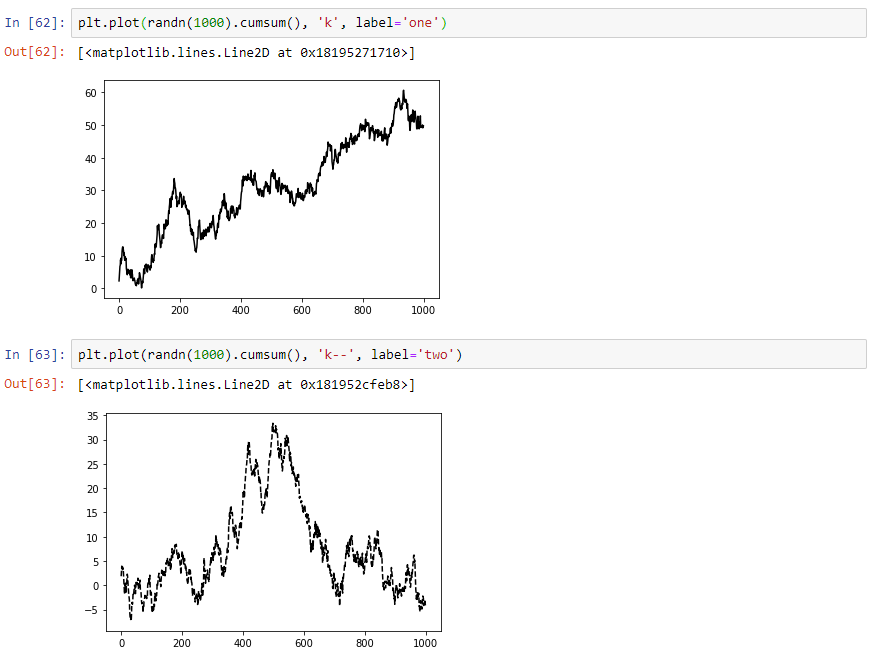


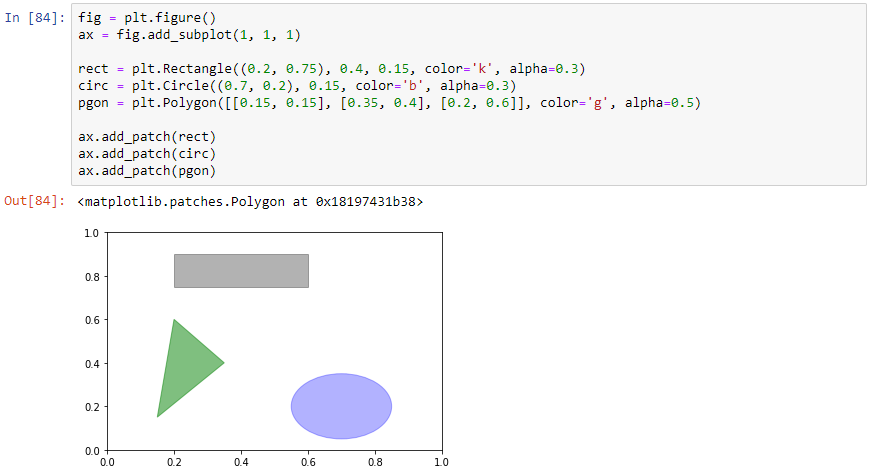
**Ticks, Labels, ve Legends**



**Legends Ekleme**





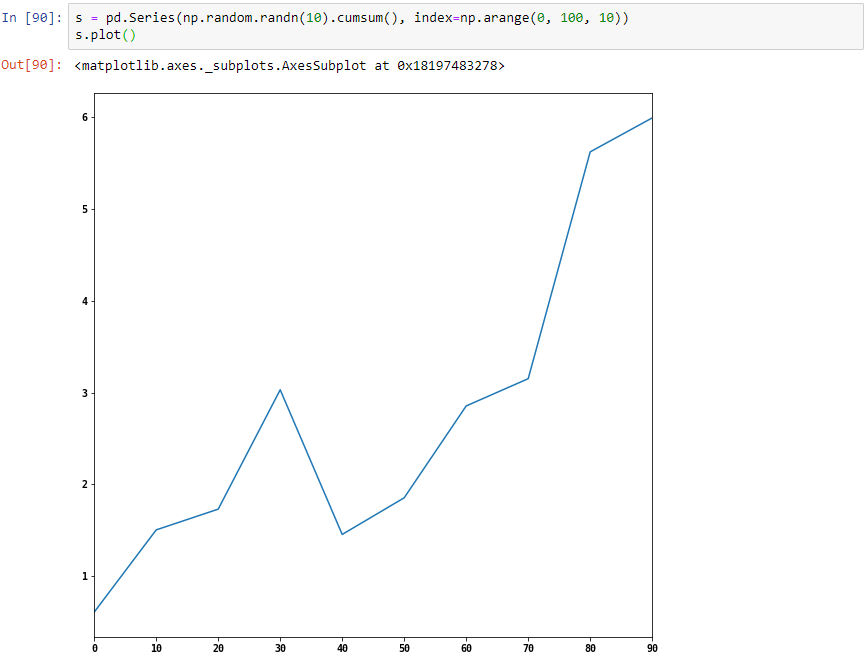


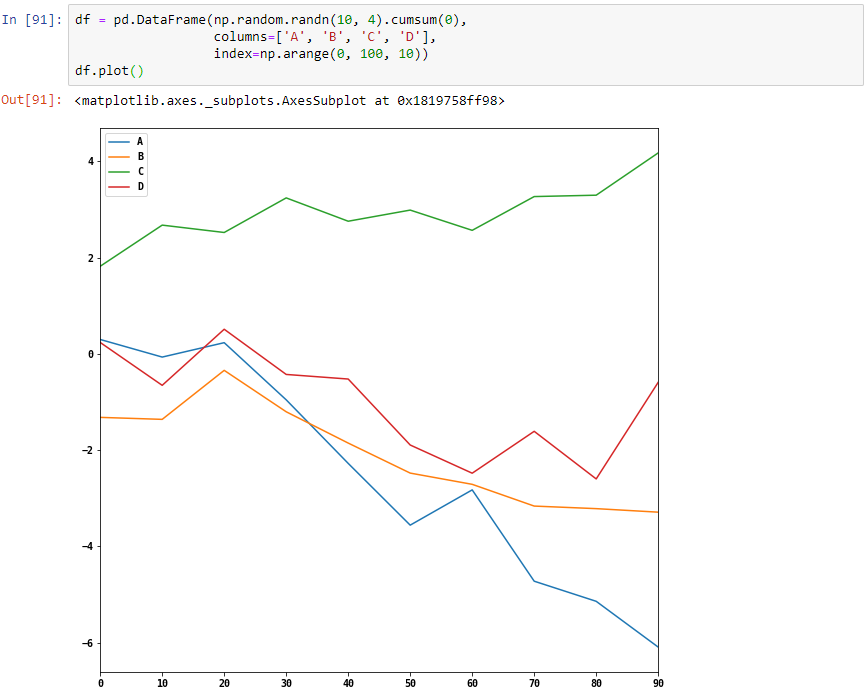
**Çizimleri Kaydetme**



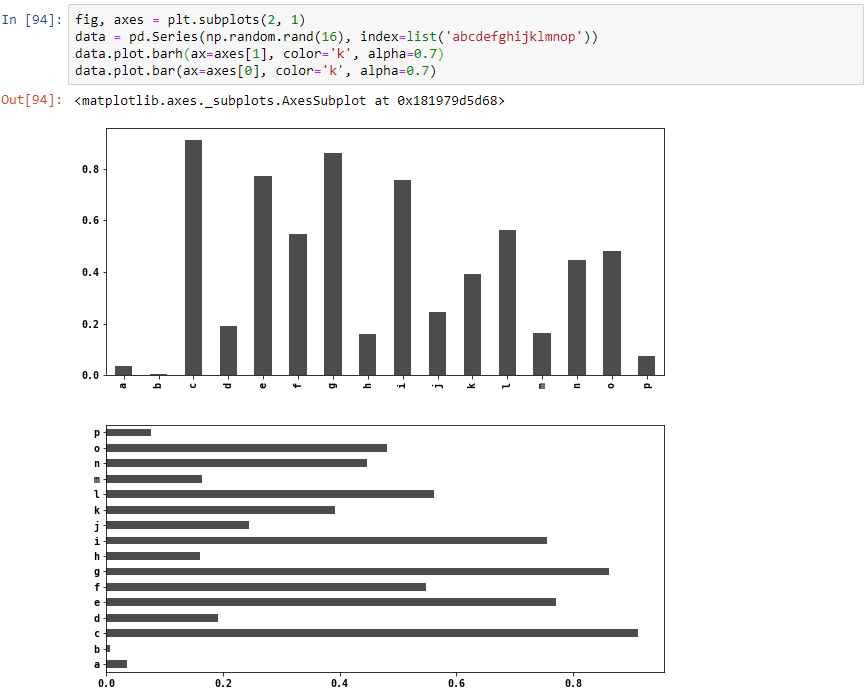
**9.2 Pandas ve Seaborn ile Çizim**

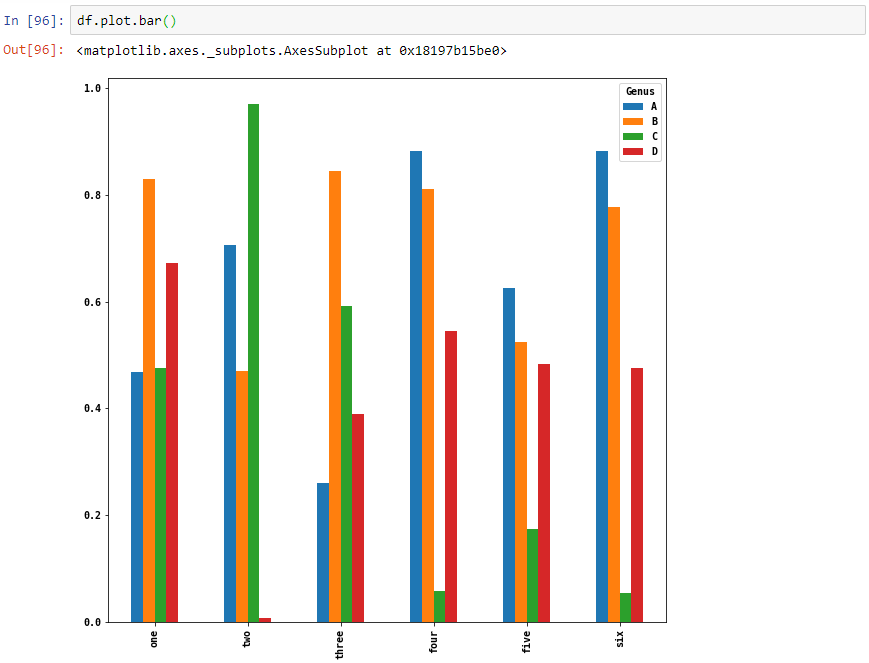
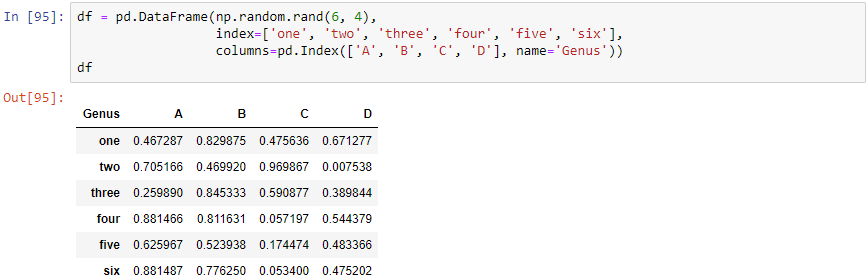
**Çizgi Gösterimi**

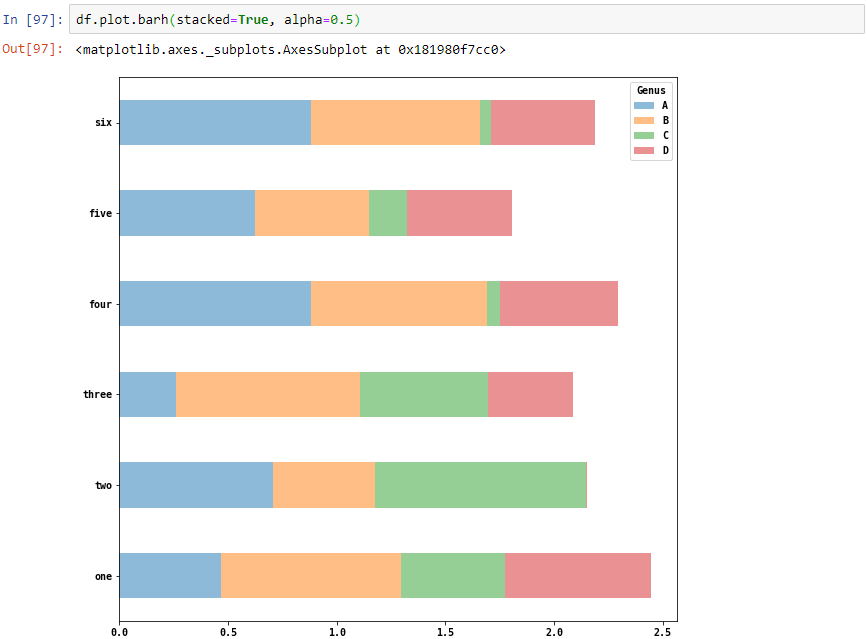


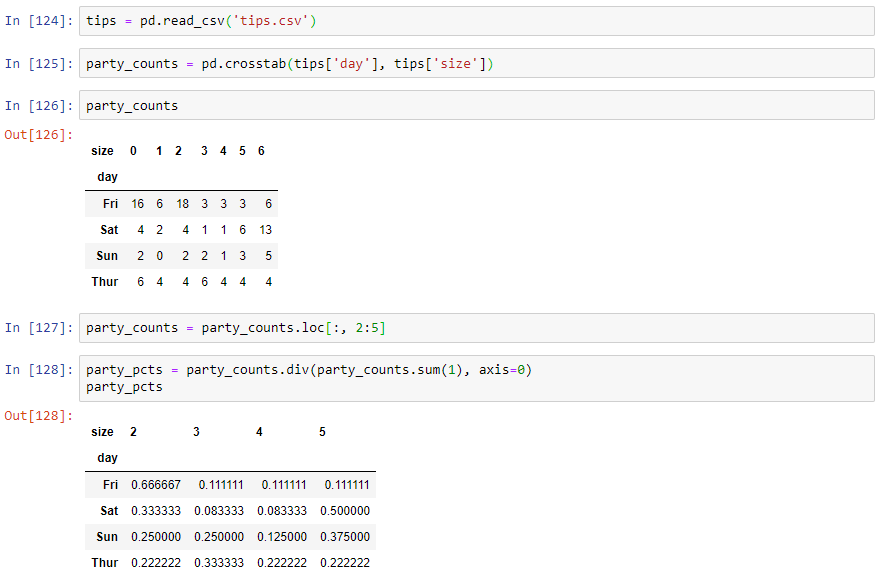


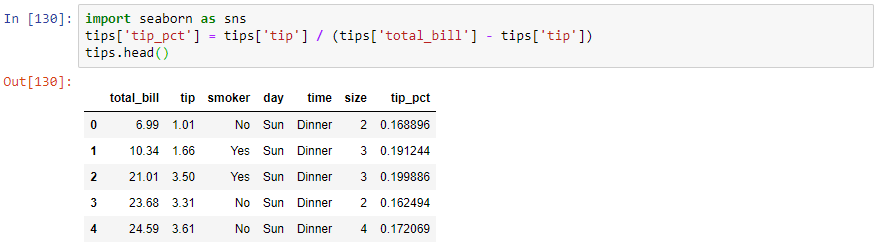
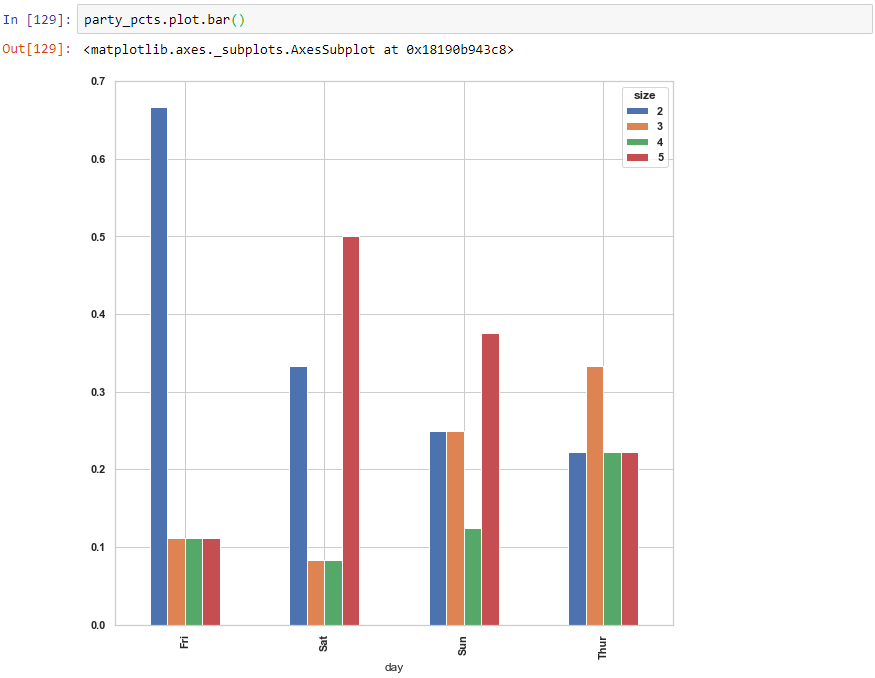
**Bar Gösterimi**

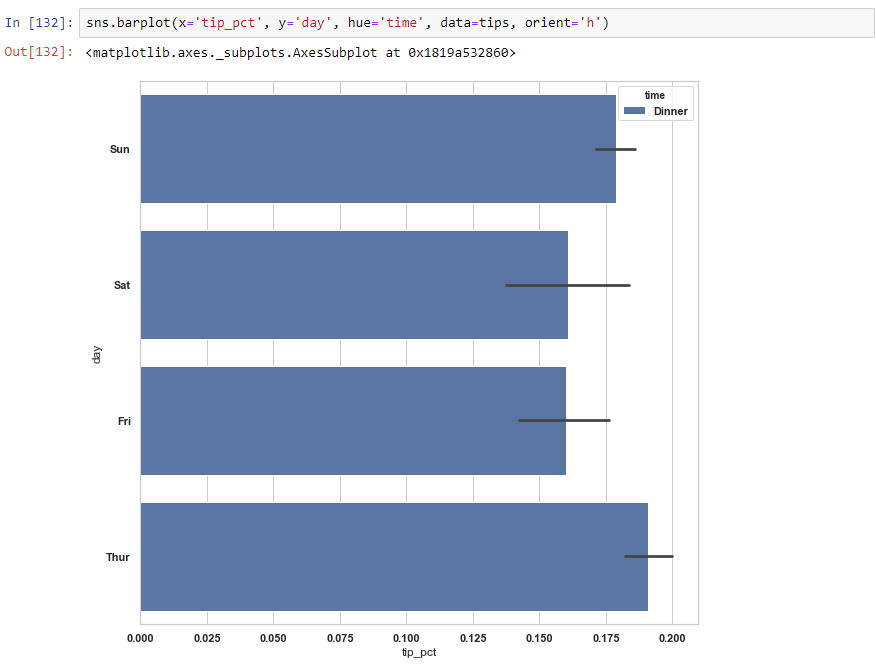
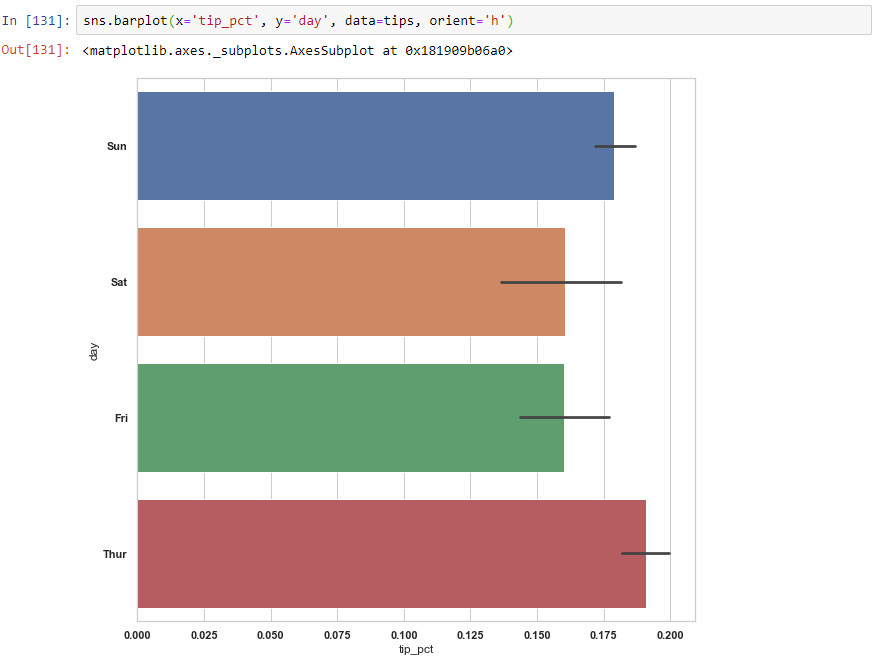




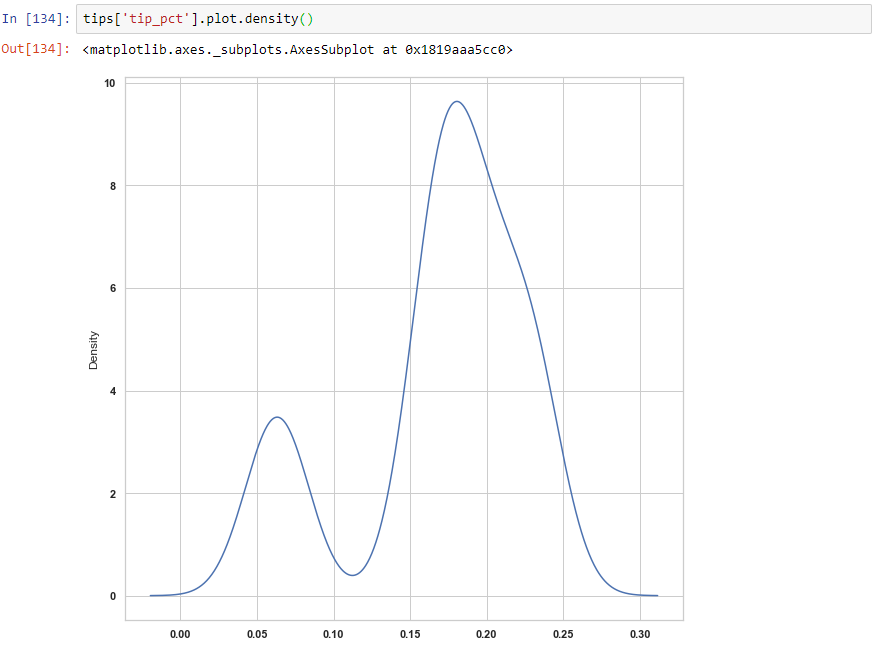
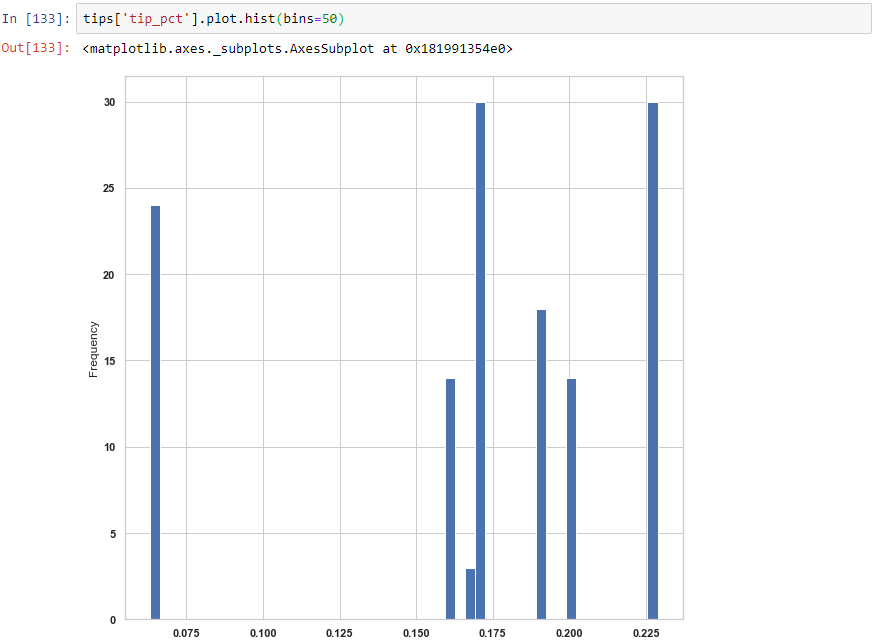


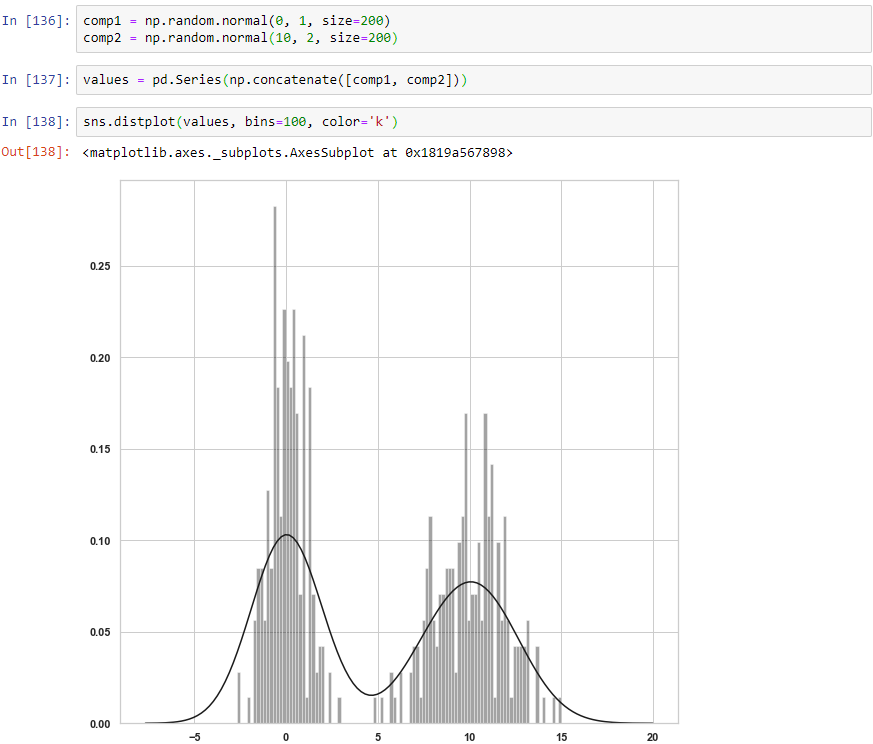




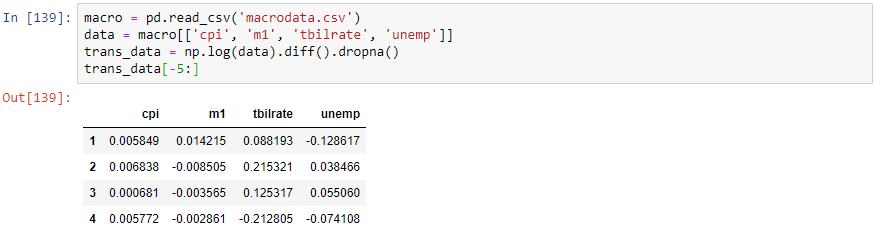


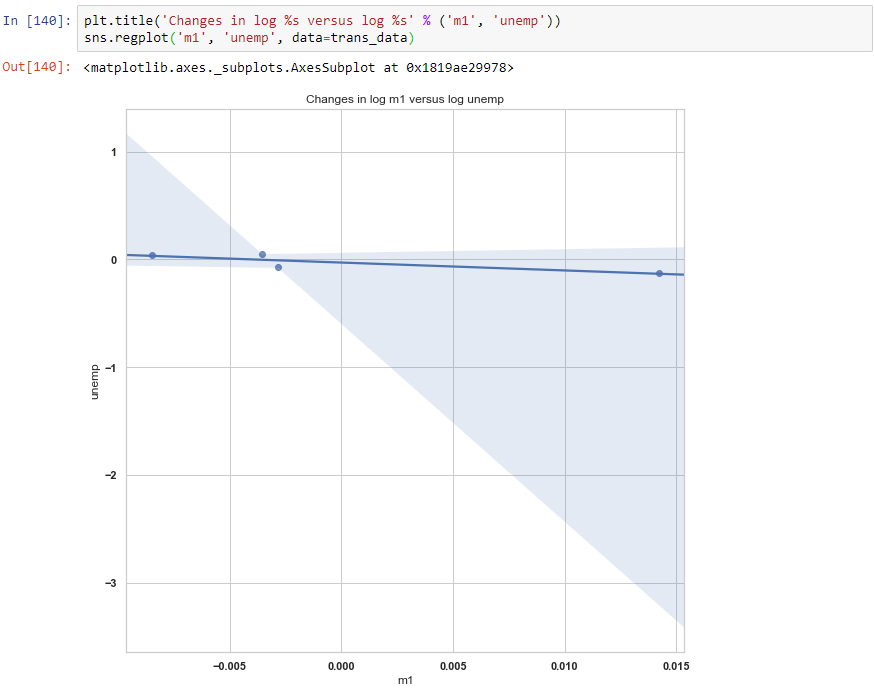
**Histogramlar ve Yoğunluk Grafikleri**





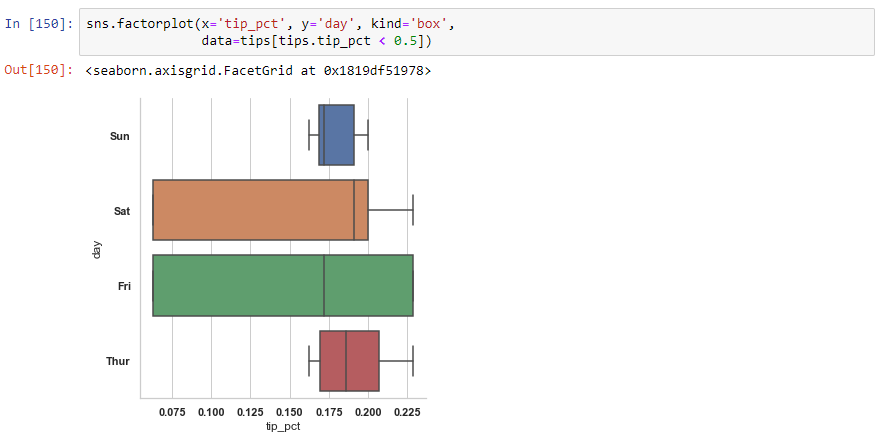
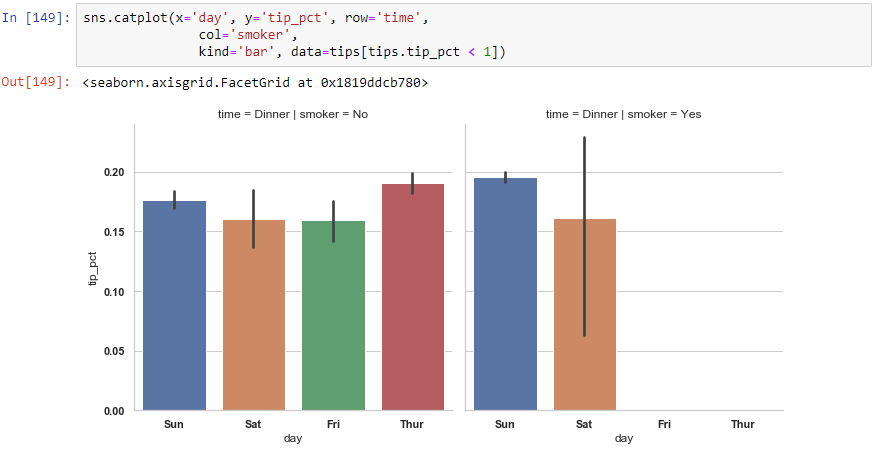
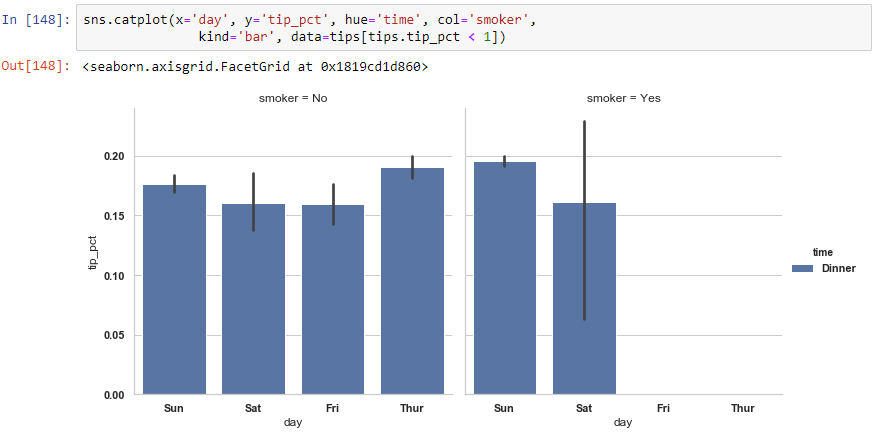
**Saçılma ve Nokta Çizgileri**







**Kategorik Veriler**

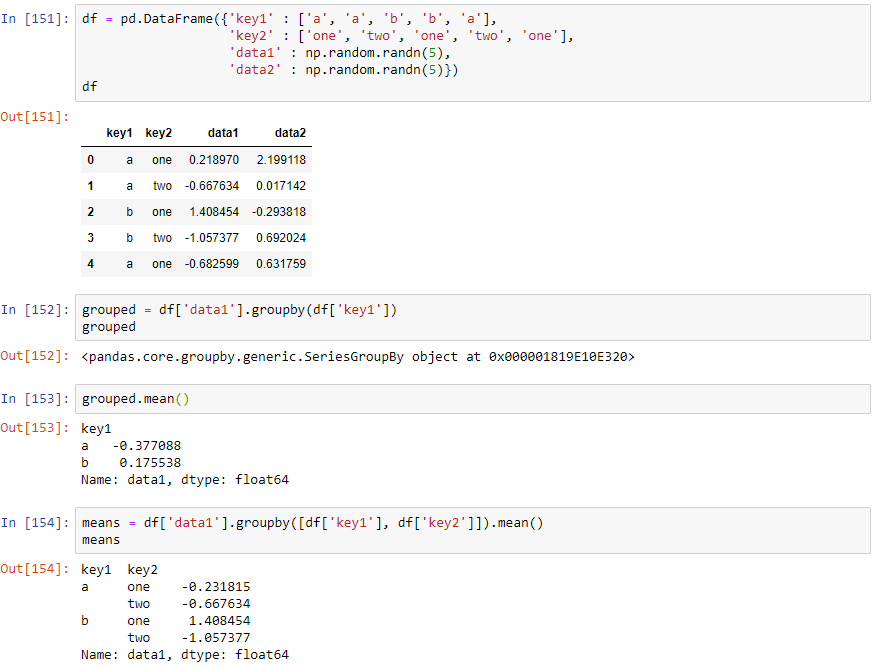


**BÖLÜM 10**

**Veri Toplama ve Grup İşlemleri**

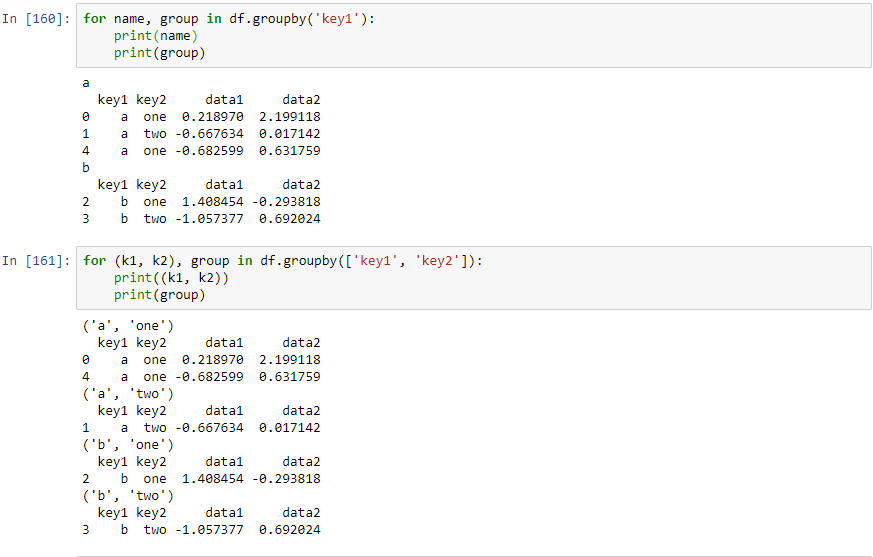
Bir veri setini kategorize etmek ve her gruba bir fonksiyon uygulamak (ister bir toplama isterse dönüşüm olsun) bir veri analizi iş akışının kritik bir bileşenidir. Bir veri setini yükledikten, birleştirdikten ve hazırladıktan sonra, raporlama veya görselleştirme amacıyla grup istatistiklerini hesaplamanız veya tabloları döndürmeniz gerekebilir. Pandas, veri kümelerini doğal bir şekilde dilimlemenize, kesmenize ve özetlemenize olanak sağlayan esnek bir grup arayüzü sunar.

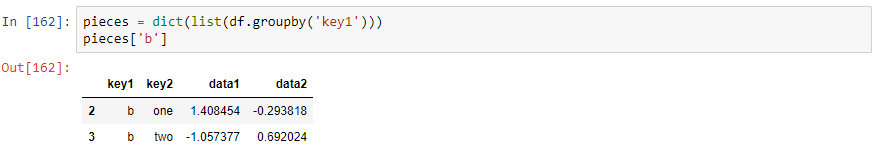
**10.1 GrupBy Mekaniği**

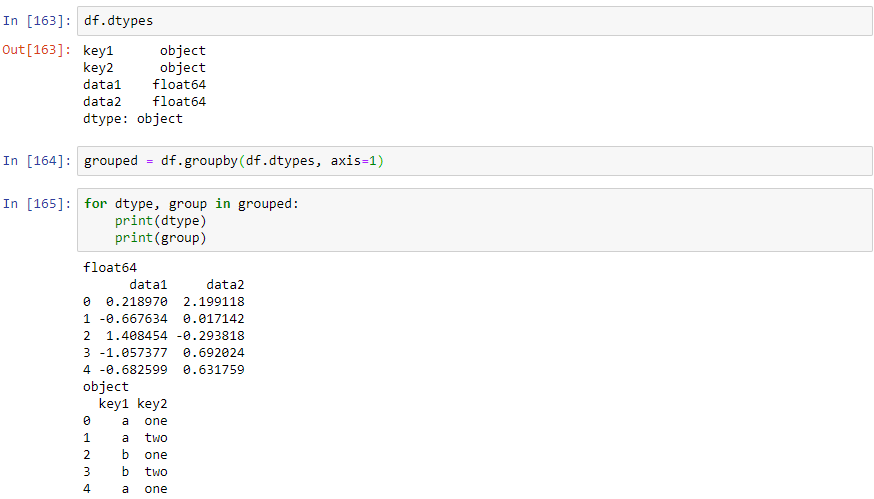




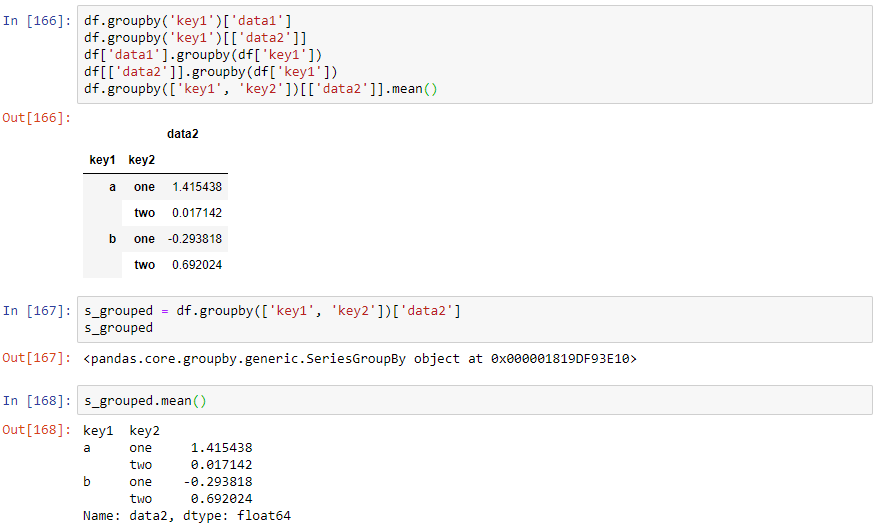
**Gruplar Üzerinde Yineleme**



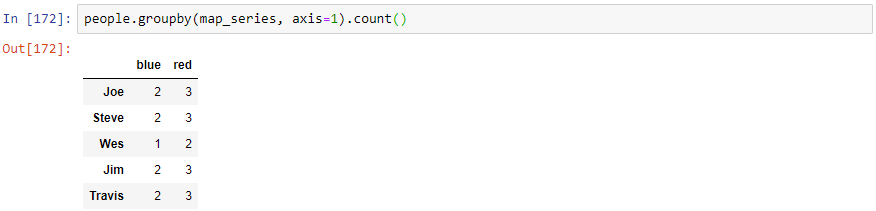
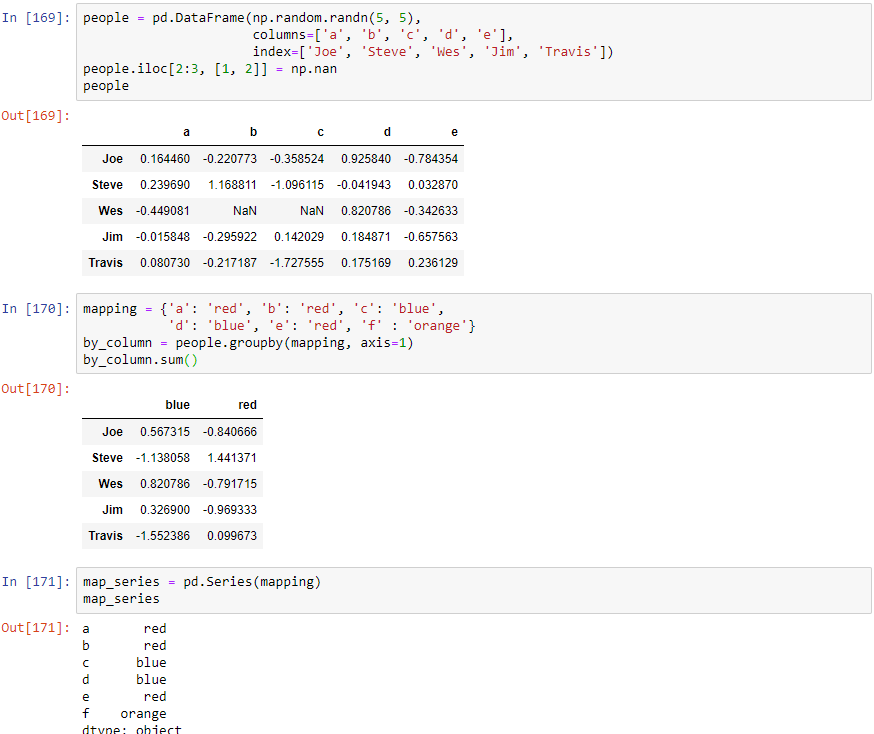




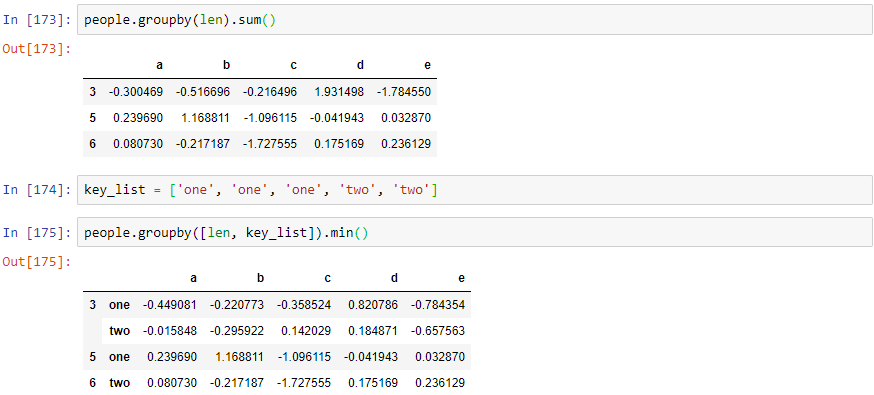
**Bir Sütun veya Sütun Alt Kümesi Seçme**



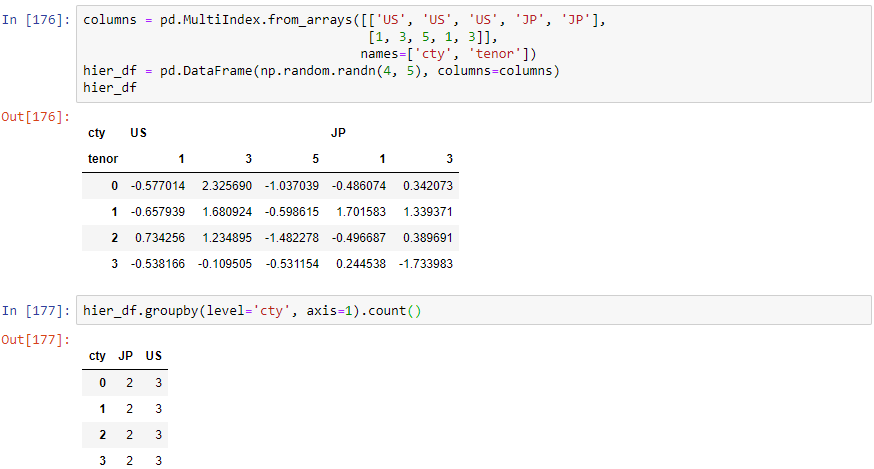
**Sözlük ve Serilerle Gruplama**



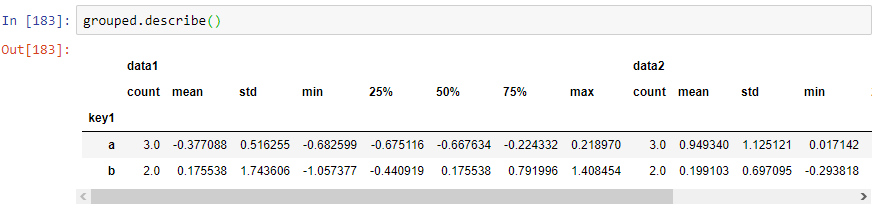
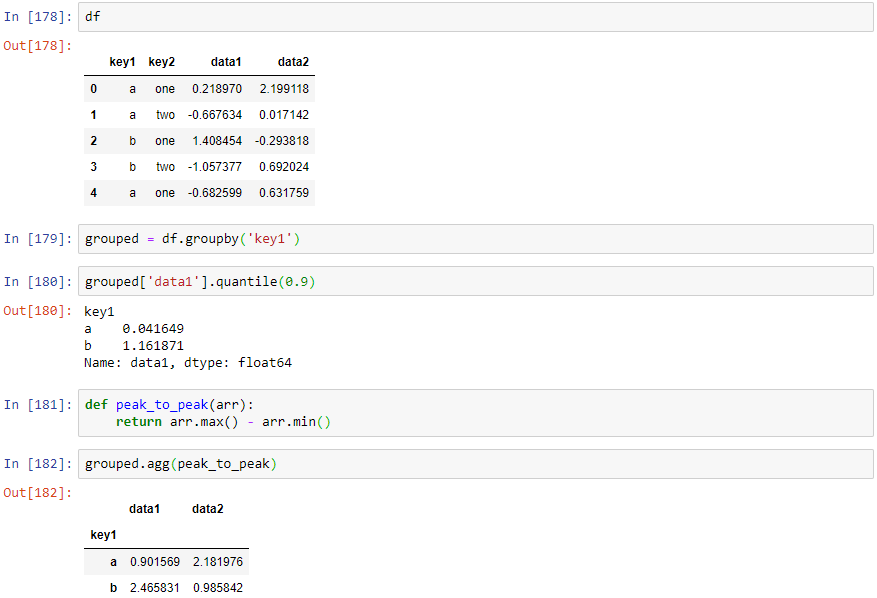
**Fonksiyonlarla Gruplandırma**



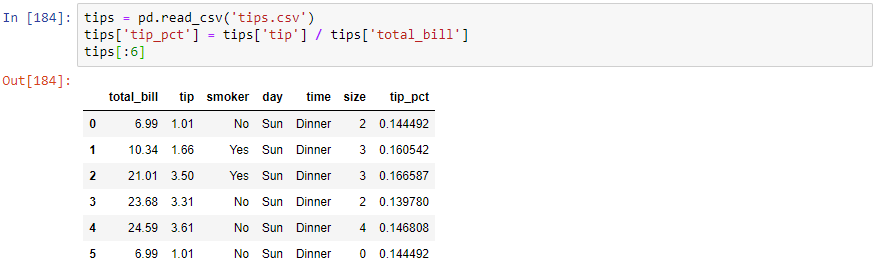
**İndeks Seviyelerine Göre Gruplandırma**

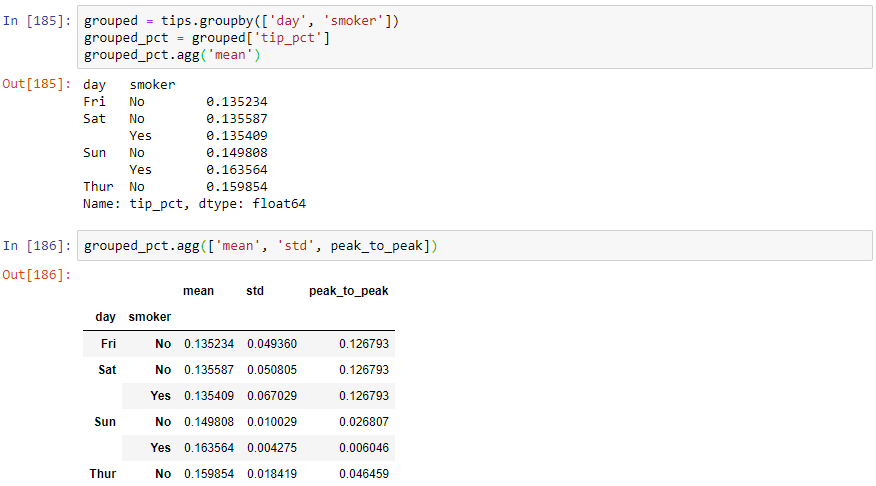


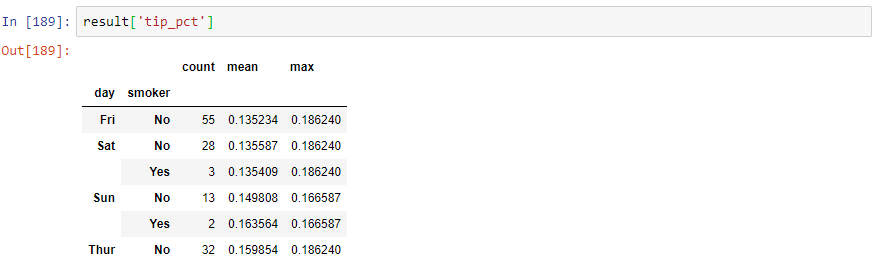
**10.2 Veri Toplama**

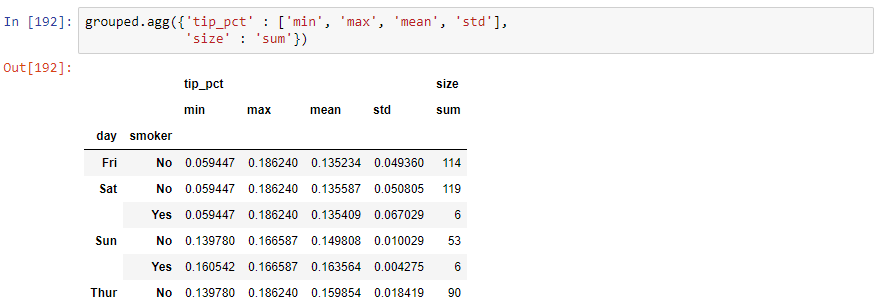
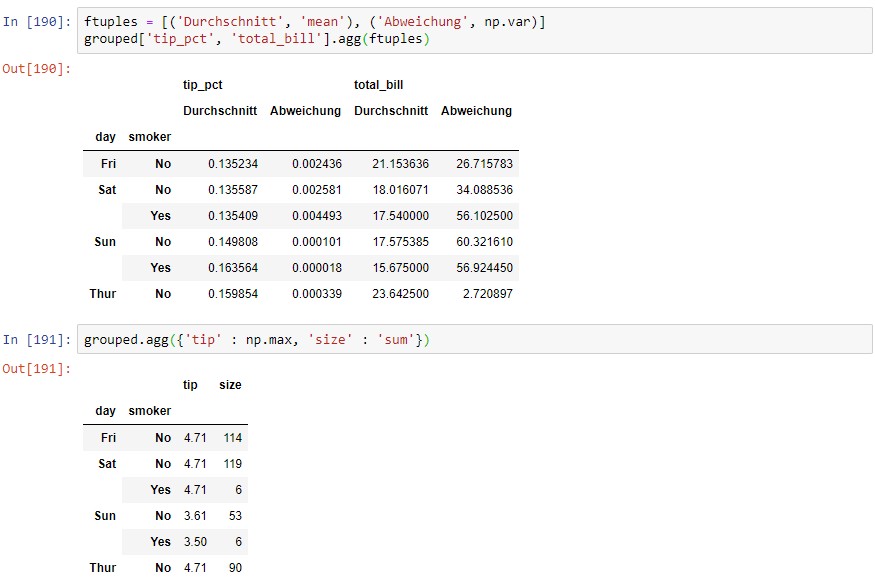


**Çoklu Fonksiyon Uygulaması**

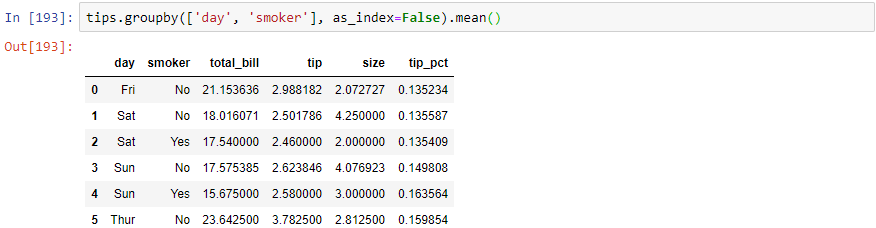




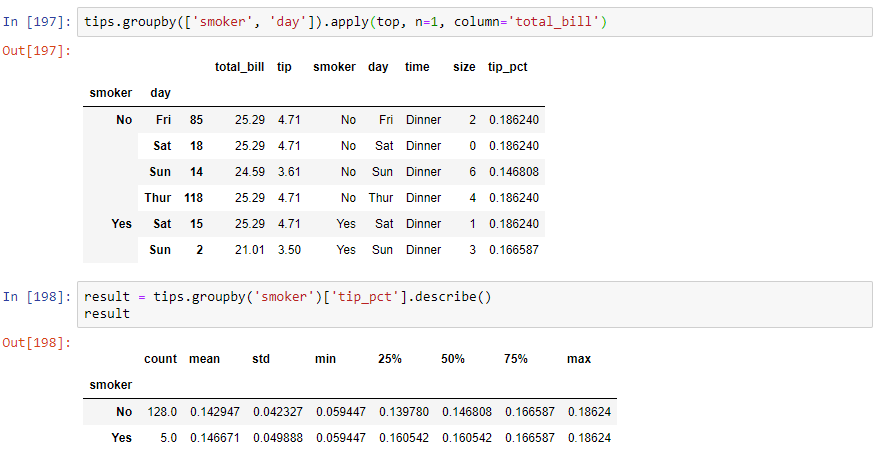
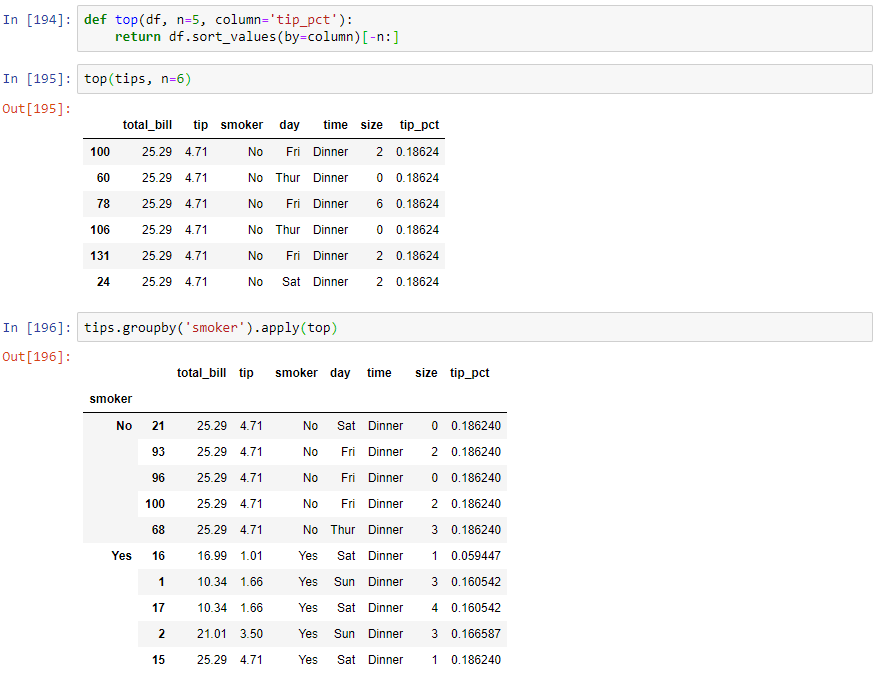


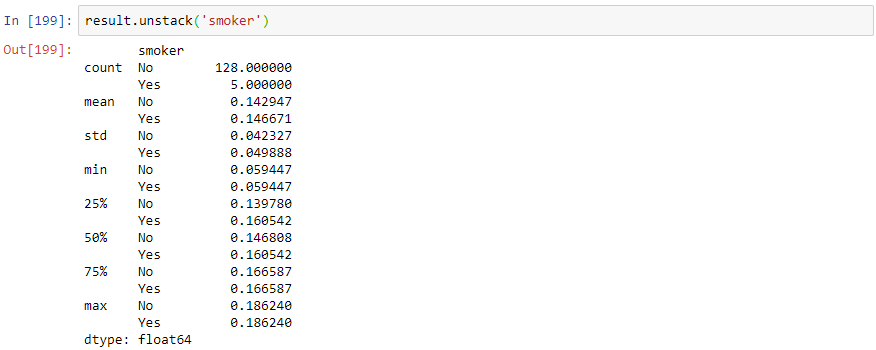


**Satır İndeksleri Olmadan Toplu Verileri Döndürme**

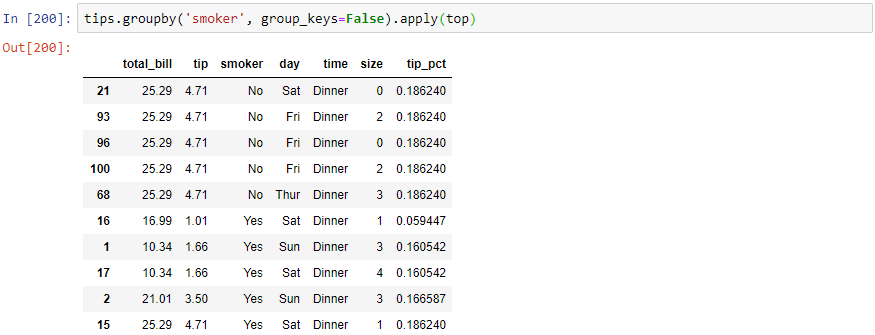


**10.3 Uygula: Split-Apply-Combine**

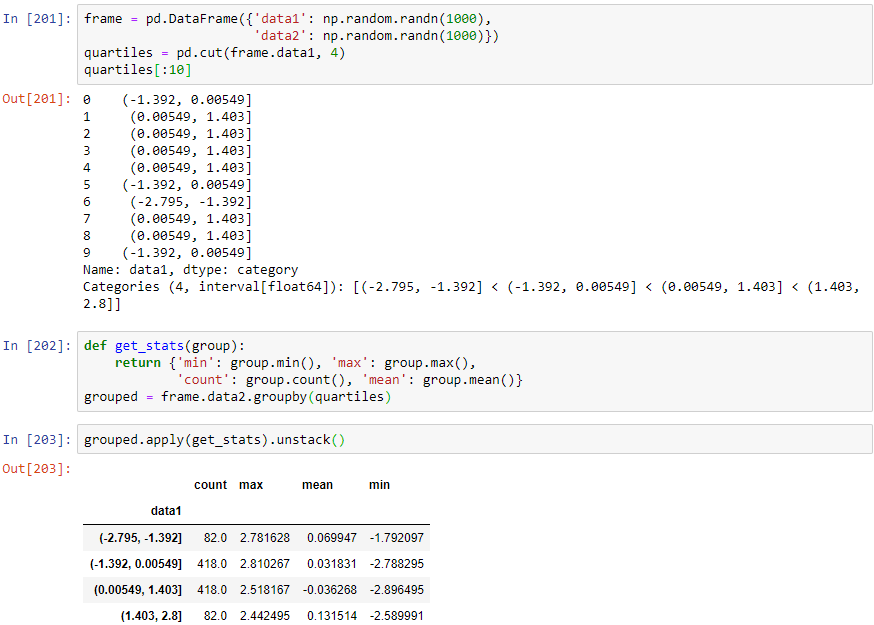


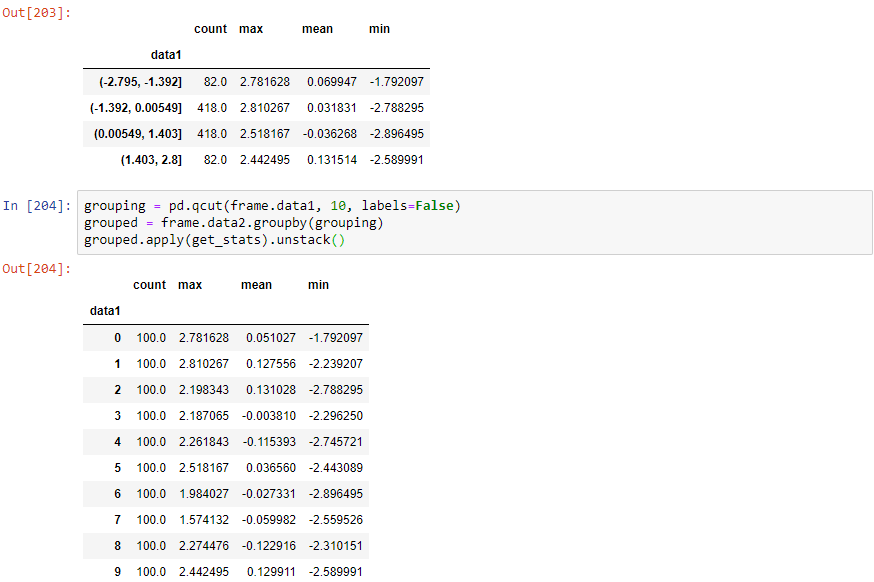


**Grup Anahtarlarını Bastırma**

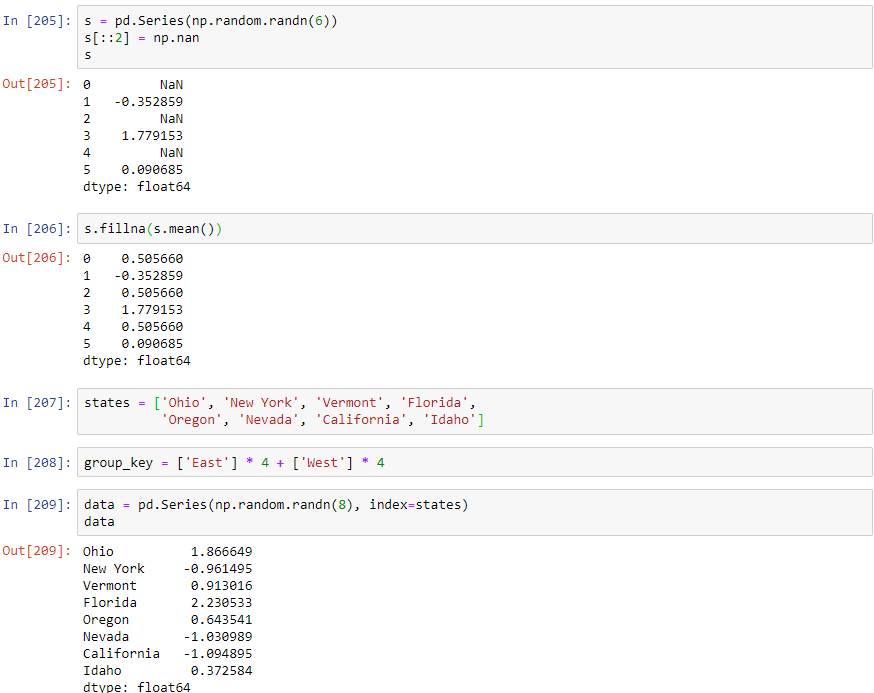


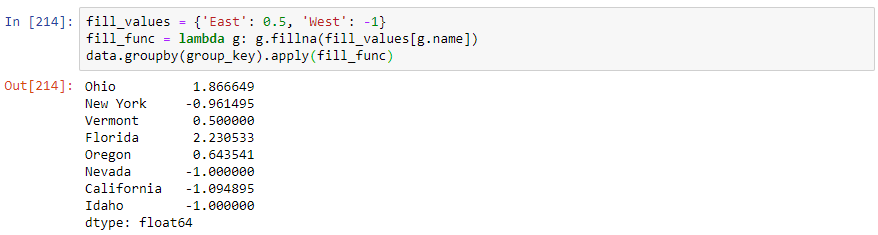
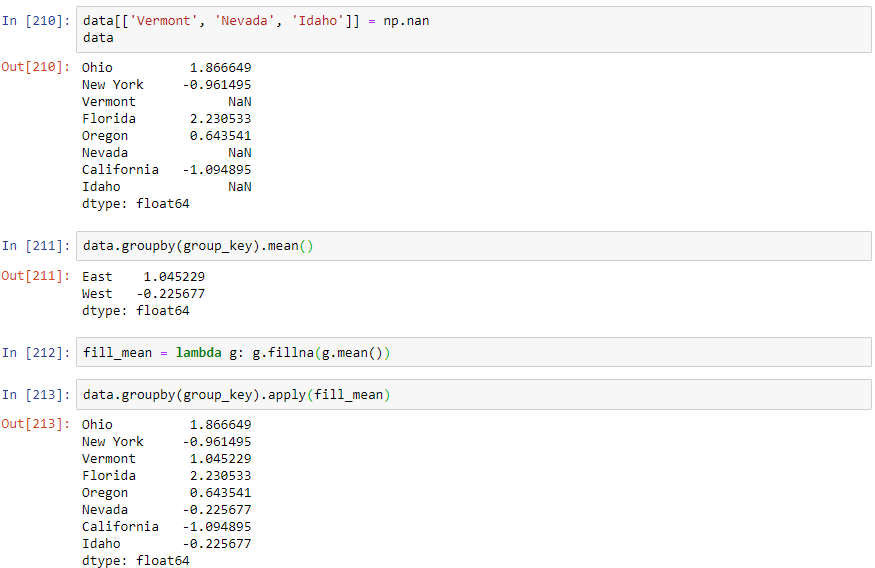
**Quantile ve Bucket Analizi**



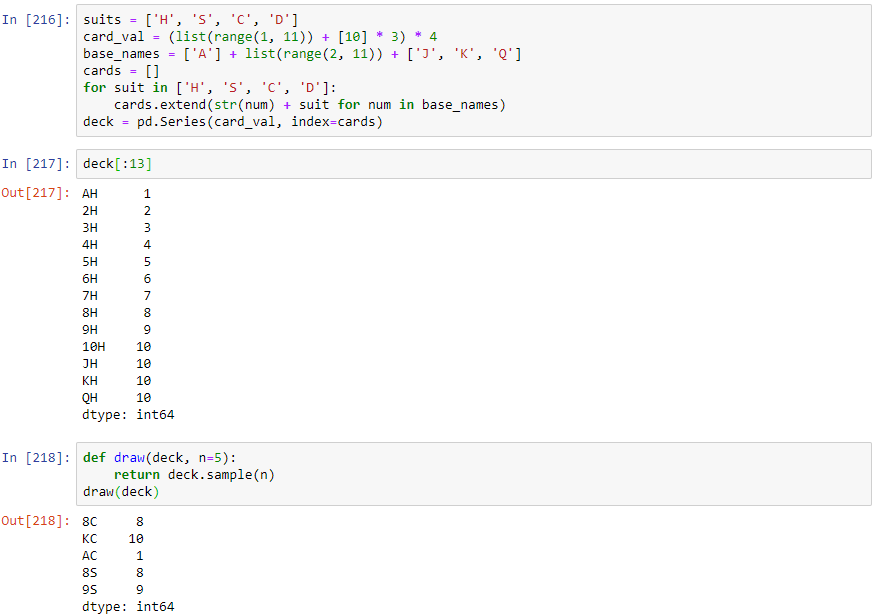


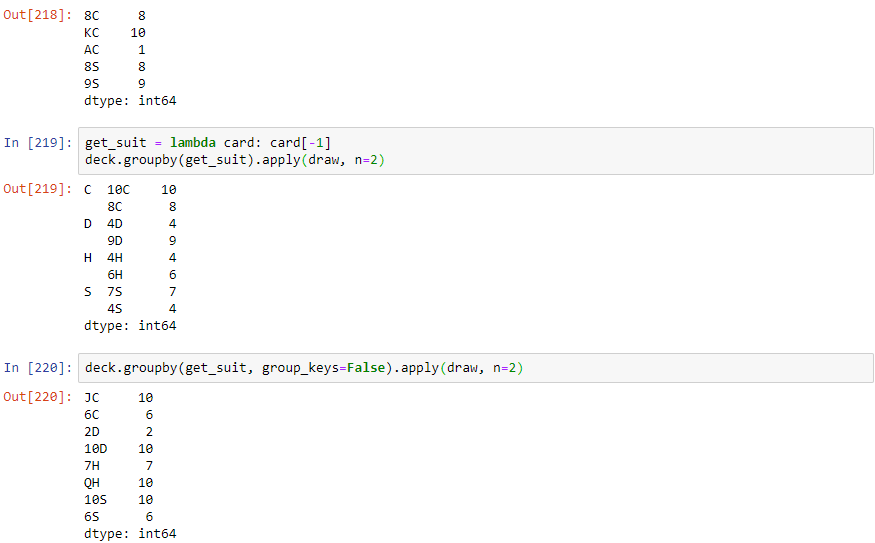
**Örnek: Eksik Değerlerin Gruba Özel Değerlerle Doldurulması**





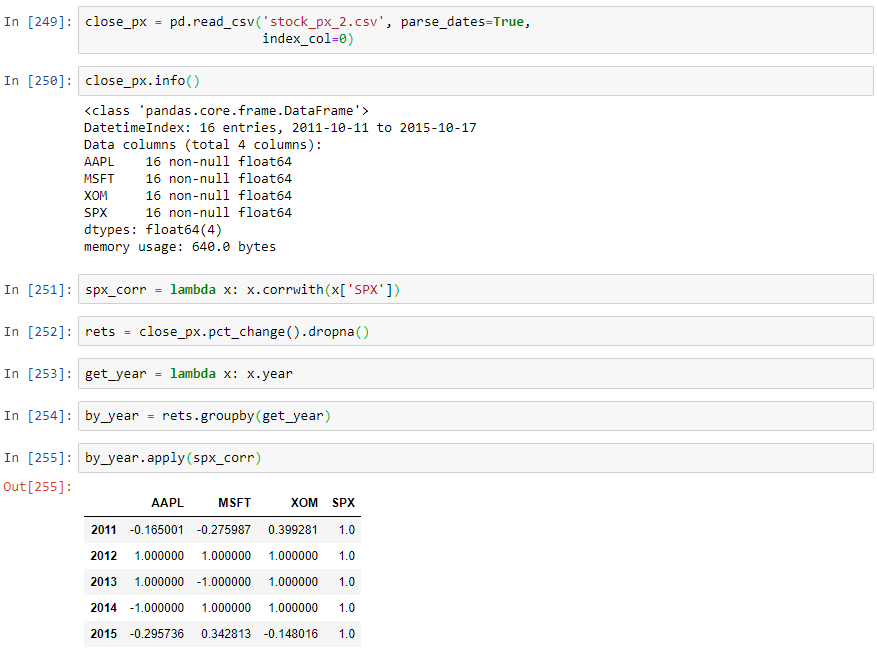
**Örnek: Rasgele Örnekleme ve Permütasyon**



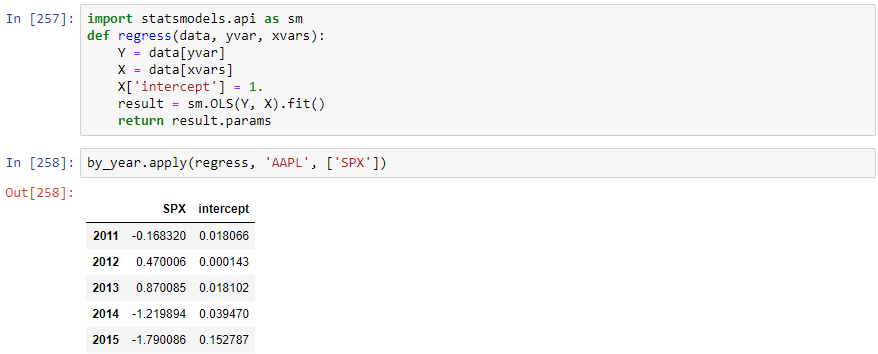


**Örnek: Grup Ağırlıklı Ortalama ve Korelasyon**

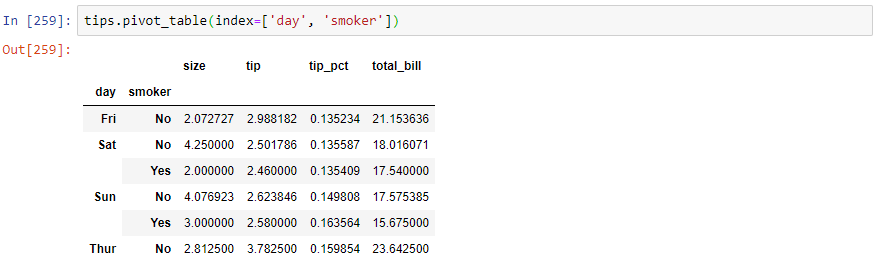


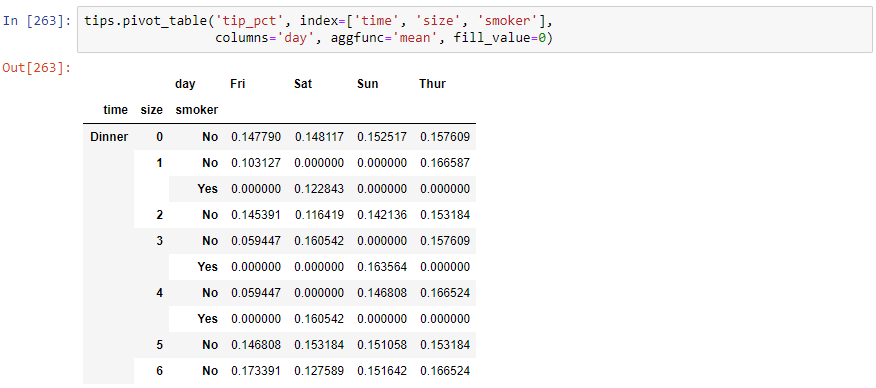
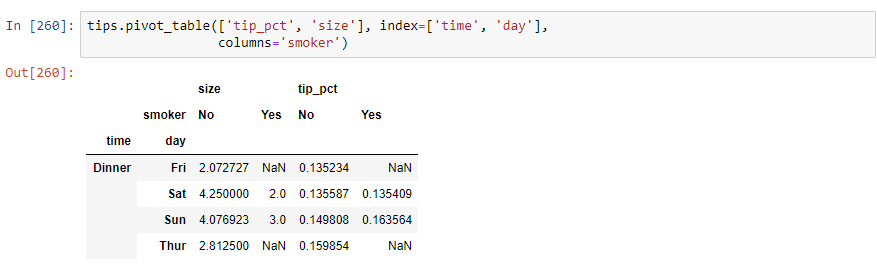


**Örnek: Doğrusal Regresyon**



**10.4 Pivot Tablolar ve Çapraz Tablolama**





**Çapraz Tablolar: Crosstab**

