

# Web Kazıma ile Rakip ve Fiyat Analizi

[illegible]





# İş Problemi

MIUUL™

www.miuul.com  
Copyright © Miuul, Inc. All Rights Reserved

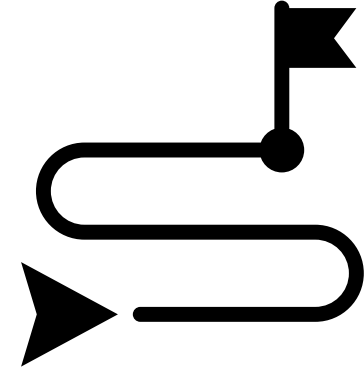
Online kitap satışı yapan bir şirket, “**Seyahat**” ve “**Kurgu Dışı**” kategorilerinde az satış yaptığını tespit ediyor.

Bu sebeple şirketin, rakip firmanın kazınması izin verilen <https://books.toscrape.com/> adlı web sitesinden “**Travel**” ve “**Nonfiction**” kategorisindeki kitaplara ait bilgileri alıp rakip ve fiyat analizi yapmaya ihtiyacı var.

Şirket sizden, bu kategorilerdeki her bir kitaba ait detay sayfalarına gitmenizi ve orada bulunan bazı bilgileri almanızı bekliyor.

miuul





# Proje Akışı

1. "Travel" ve "Nonfiction" kategorilerine ait kitapların yer aldığı sayfa linkleri ana sayfa içerisinden alınmalı.
2. Sonrasında ilk kategoriye ait ürünlerin bulunduğu sayfa görüntülenerek, tüm kitapların detay sayfalarına erişmek için linkler kazınmalı.
3. Kazınan linkler kullanılarak o kategoriye ait tüm kitapların detay bilgileri kazınmalı.
4. Adım 2 ve 3'ü diğer kategori için de yapabilmek adına sistem otomatize edilmeli.



# Proje Görevleri

---

## Görev 1: Tarayıcıyı Konfigüre Etme ve Başlatma

---

**Adım 1:** Webdriver sınıfını kullanarak bir adet options adında ChromeOptions tanımlayınız.

**Adım 2:** Tanımladığınız options'a tam ekran özelliği ekleyiniz.

**Adım 3:** Bir önceki adımlarda hazırladığınız options'ı da kullanarak bir adet driver adında Chrome tarayıcısı oluşturunuz.

## Görev 2: Ana Sayfanın İncelenmesi ve Kazınması

---

**Adım 1:** Ana Sayfayı driver ile açınız ve inceleyiniz.

**Adım 2:** "Travel" ile "Nonfiction" kategori sayfalarının linklerini tutan elementleri tek seferde bulan XPath sorgusunu yazınız.

**Adım 3:** XPath sorgusu ile yakaladığınız elementleri driver'ı kullanarak bulunuz ve kategori detay linklerini kazıyınız.

## Görev 3: Kategori Sayfasının İncelenmesi ve Kazınması

---

**Adım 1:** Herhangi bir detay sayfasına girip o sayfadaki kitapların detay linkini tutan elementleri yakalayan XPath sorgusunu yazınız.

**Adım 2:** driver ile XPath sorgusunu kullanarak elementleri yakalayınız ve detay linklerini çıkarınız.

- 
- 
- 
- 
-



## Görev 3: Kategori Sayfasının İncelenmesi ve Kazınması

•  
•  
•

**Adım 3:** Sayfalandırma (Pagination) için butonlara tıklamak yerine sayfa linkini manipüle etme yöntemini kullanınız ve döngü kullanarak sayfalandırma sistemi ekleyiniz.

**Adım 4:** Sayfalandırmanın sonuna geldiğinizi anlamak adına kategorinin 999. sayfasına giderek karşınıza çıkan sayfayı inceleyiniz.

İpucu: ....category/books/nonfiction\_13/page-999.html

**Adım 5:** Bir önceki adımdaki incelemenin sonucunda sayfalandırmada sonsuz döngüye girmemek adına bu durumu kontrol ediniz.

İpucu: text'inde 404 içeren bir h1 var mı?) veya (if not product\_list) ya da (len(product\_list) <= 0

## Görev 4: Ürün Detay Sayfasının Kazınması

**Adım 1:** Herhangi bir ürünün detay sayfasına girip class attribute'ı content olan div elementini yakalayınız.

**Adım 2:** Yakaladığınız divin html'ini alınız ve inner\_html adlı değişkene atayınız.

**Adım 3:** inner\_html ile soup objesi oluşturunuz.

**Adım 4:** Oluşturduğunuz soup objesi ile şu bilgileri kazıyınız:

- Kitap Adı
- Kitap Fiyatı
- Kitap Yıldız Sayısı
- Kitap Açıklaması
- Product Information Başlığı altında kalan tablodaki bilgiler.



## Görev 5: Fonksiyonlaştırma ve Tüm Süreci Otomatize Etme

**Adım 1:** İşlemleri fonksiyonlaştırınız.

Örnek olarak:

- `def get_product_detail(driver)`
- `def get_category_detail_urls(driver)`

**Adım 2:** Süreci otomatize ederek, Travel ile Nonfiction kategorilerine ait tüm kitapların detaylarını alacak şekilde kodu düzenleyiniz.

**İpucu:** For döngüleri kullanabilirsiniz.

**miuul**

miuul.com