

Data Science Salaries 2020-2023

Giriş:

Bu veri kümesinin, veri bilimi maaşlarını etkileyen çeşitli faktörleri anlamak ve analiz etmenin yanı sıra bu meslekte çalışan bireylerin maaşlarını tahmin etmek veya tahmin etmek amacıyla oluşturulduğu görülmektedir. Veriler muhtemelen birden fazla web sitesindeki iş ilanlarının toplanmasıyla toplanmıştır. Bu veri kümesinin temel amacı, veri bilimi maaşlarını etkileyen faktörlere ilişkin değerli bilgiler sağlamaktır. Bu veri kümesi ve tahmin görevinin amacı, veri bilimi alanına girmeyi hedefleyen iş arayanlara yardımcı olmaktır. Deneyim, konum ve iş rolü gibi faktörlere dayalı olarak çeşitli veri bilimi rollerindeki maaşları tahmin eden bir makine öğrenimi modeli geliştirerek, iş arayanlara değerli bilgiler sağlamayı amaçlıyorum.

Veri Seti Tanımı:

Bu veri seti, farklı iş pozisyonlarına ve istihdam türlerine sahip çalışanların 2020 ve 2023 yılları arasındaki deneyim düzeylerini, maaşlarını ve diğer ilgili bilgileri içeren bir kaynaktır. Veri seti, iş yılı (work_year), deneyim seviyesi (experience_level), istihdam türü (employment_type), iş başlığı (job_title), maaş (salary), maaş para birimi (salary_currency), ABD doları cinsinden maaş (salary_in_usd), çalışanın ikamet ettiği yer (employee_residence), uzaktan çalışma oranı (remote_ratio), şirket konumu (company_location) ve şirket büyüklüğü (company_size) gibi çeşitli özellikleri içermektedir.

Veri seti, farklı ülkelerdeki iş pozisyonları ve şirketlerin büyüklükleri hakkında bilgi sunarak çalışan maaşları ve deneyim seviyeleri arasındaki ilişkileri incelemek için kullanılabilecek kapsamlı bir veri kaynağıdır. Bu veri seti, veri bilimi, istihdam ve maaş analizi gibi bir dizi konuda araştırma ve analiz için potansiyel sunmaktadır.

Veri Keşfi ve Temizleme:

İlk olarak, veri setindeki bazı sütunlardaki kısaltmalar açıklayıcı terimlere dönüştürüldü. Bu dönüşümler, veri setinin anlaşılabilirliğini artırdı. Örneğin, "SE" (Senior) olarak dönüştürüldü. Analizin amacına uygun olmayan 'salary' ve 'salary_currency' sütunları, maaş verilerinin işlenmediği bu analizden çıkarıldı. Veri setinin daha iyi anlaşılması için çeşitli görselleştirmeler oluşturuldu. Bu görselleştirmeler aşağıdaki sonuçları gösterdi: Deneyim seviyeleri (experience_level), istihdam türleri (employment_type), ve şirket büyüklükleri (company_size) gibi kategorik özelliklerin dağılımları grafiklerle gösterildi. Maaşların dağılımı, histogramlar ve kutu grafikleri ile analiz edildi. Uzaktan çalışma oranlarının dağılımı görselleştirildi.

Veri Analizi:

Veri setinin temel özet istatistikleri hesaplandı. Bu istatistikler, iş yılı (work_year), maaş (salary_in_usd) ve uzaktan çalışma oranı (remote_ratio) gibi önemli özellikler için aşağıda verilen sonuçları sağladı: İş yılı (work_year) için ortalama 2022.37, standart sapma 0.69 ve minimum 2020'de, maksimum ise 2023'te bulunuyor. Maaş (salary_in_usd) için ortalama 190,695 dolar, standart sapma 671,676 dolar ve minimum 6,000 dolar ile maksimum 30,400,000 dolar arasında değişiyor. Uzaktan çalışma oranı (remote_ratio) için ortalama olarak %46.27, standart sapma %48.59 ve minimum %0, maksimum ise %100 olarak görülmekte. Veri setindeki maaşların analizi yapıldı. Çoğu maaşın 100,000 ile 200,000 dolar arasında yoğunlaştığı görüldü. Ancak, yüksek maaşlı pozisyonlarda aykırı değerlerin olduğu gözlemlendi. Veri setindeki uzaktan çalışma oranları incelendi. Yaklaşık olarak çalışanların %56'sının uzaktan çalıştığı, %42'sinin ofiste çalıştığı ve %0.7'sinin hibrit bir çalışma düzeni tercih ettiği görüldü.

Sonuçlar ve Yorumlar:

Uzaktan çalışma seçeneğinin 2020'den 2023'e önemli ölçüde azalması, belki de doğrudan kaldırılan COVID-19 ile ilgili kısıtlamalarla ilgilidir. Veri setinde temsil edilen şirketin merkez ofisinin ve çalışanlarının ikametgahının ezici çoğunluğu ABD'dir. Büyük ve orta ölçekli şirketlerde ortalama maaş 2020-2023 yılları arasında önemli ölçüde artarken, küçük şirketler aynı dönemde oldukça istikrarlı bir ortalama maaş seviyesi gösteriyor. Ortalama en yüksek maaş orta ölçekli şirketlerde dir. Tam zamanlı istihdam türünde beklenen ortalama maaş diğerlerine göre daha yüksektir çünkü üst düzey pozisyonlar için maaş seviyelerinde büyük bir fark var. Modelleme kısmında genel olarak kullanılan üç modelin tümü iyi çalışmadı ve ek ayarlama veya başka modellerinde test edilmesini gerektiriyordu.

MERT OLGUN