

Data Visualization Assignment 4

Ezgi Altıntop
41623432038

From the given 13 cases, 2. Data Scientist Salaries and 3. Top-Selling Books were selected. First, these datasets were examined and the necessary questions were answered.

1.Data Scientist Salaries

Data Source: [Data Scientist Salaries](#)

Dataset Summary:

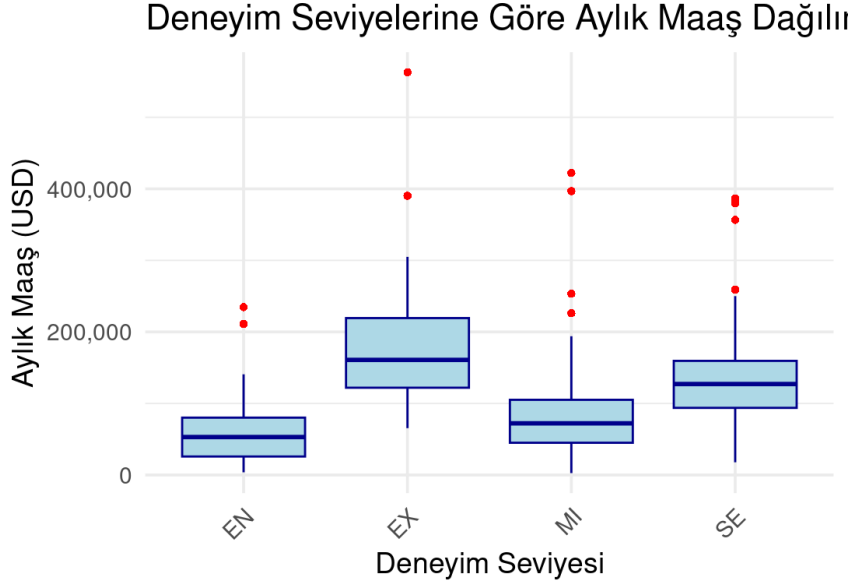
This dataset contains salary information for professionals working in various data-related roles (e.g., Data Scientist, Data Engineer, Data Analyst, Machine Learning Engineer) across different countries, experience levels, and working conditions. Each row generally represents a single role or position, including salary details and related attributes.

Key Variables:

- **work_year:** The year for which the salary information is relevant.
- **experience_level:** The employee's level of experience, typically categorized as:
 - Junior or Entry-level (e.g., 'Jr' or 'EN')
 - Mid-level (e.g., 'MI')
 - Senior-level (e.g., 'SE' or 'Sr')
 - Executive-level (often at a higher managerial or strategic position)
- **employment_type:** The type of employment (full-time, part-time, contract, freelance).
- **job_title:** The specific role (e.g., Data Scientist, Data Engineer, Data Analyst).
- **salary:** The raw salary figure, often recorded in the local currency of the role's location.
- **salary_currency:** The currency in which the salary is paid.
- **salary_in_usd:** The salary converted into US dollars, enabling easier comparison across different regions.
- **employee_residence:** The country where the employee resides.
- **remote_ratio:** The percentage of time the role is performed remotely:
 - 0: Fully on-site
 - 50: Hybrid (partially remote)
 - 100: Fully remote
- **company_location:** The country where the employing company is located.
- **company_size:** The size category of the company:
 - S: Small
 - M: Medium
 - L: Large

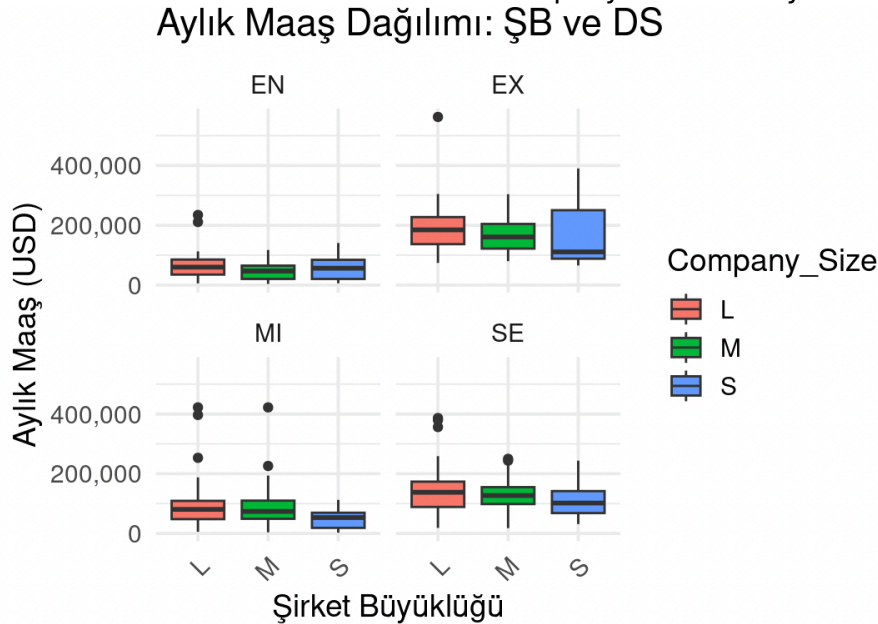
Questions:

- I. Examine the monthly salary distribution (in USD) of data science roles by experience levels.



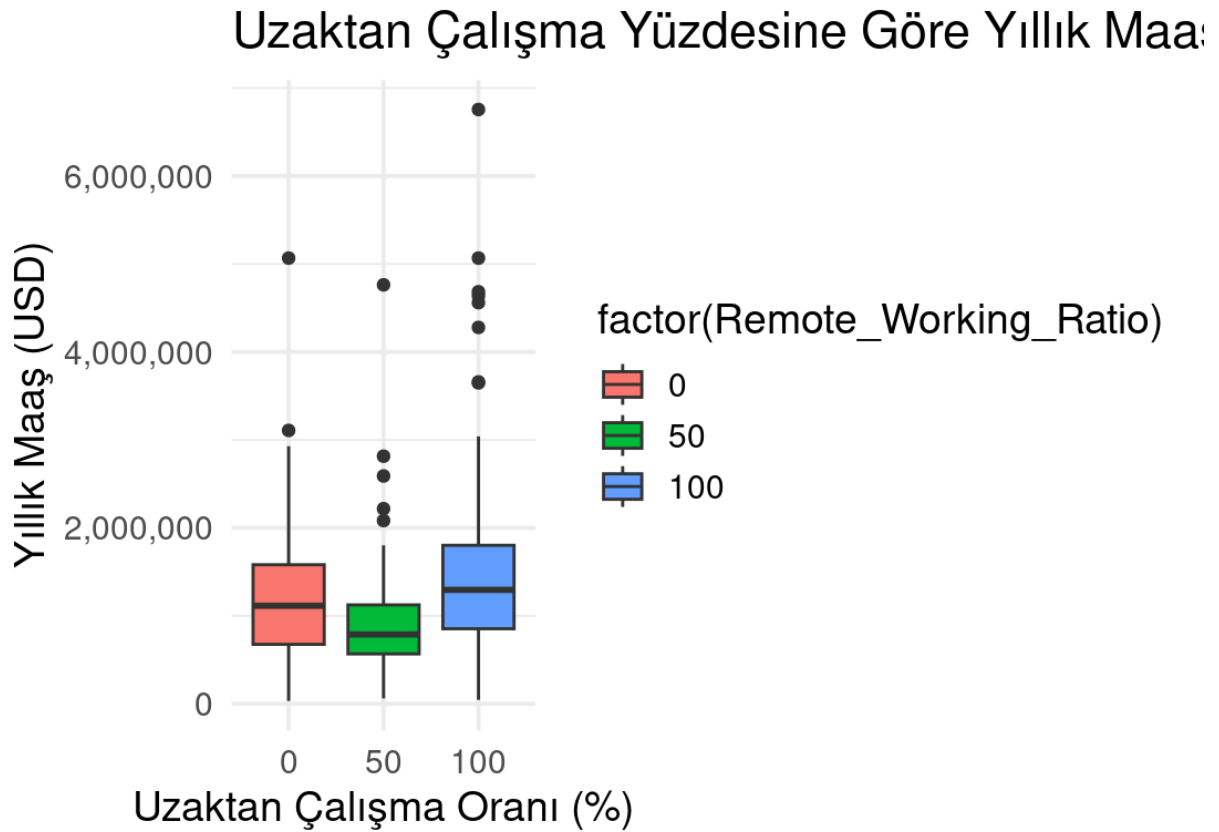
The graph shows that there is a positive relationship between experience level and salary. It is seen that the average/median salary increases steadily as the career progresses, and the salary range can widen as experience increases. This emphasizes that experience in the field of data science is an important factor in determining the salary level.

- II. Analyze the monthly salary distribution by company size and experience level. Determine the most suitable company size for entry-level candidates.



The most suitable company size for entry-level (EN) candidates is large (L) or at least medium (M) companies in terms of salary. Median salaries for entry-level positions in smaller companies are lower than in these two categories. Therefore, it seems more advantageous for an entry-level candidate to look to large or medium-sized companies when it comes to salary expectations.

- III. Investigate the variation in annual salary distribution based on the remote work percentage system.



The graph shows that as the rate of remote work increases (especially when it reaches 100%), the median annual salary increases and the salary range widens. This may suggest that fully remote positions may be supported by more competitive salaries or that such positions attract more qualified and therefore higher paid employees. However, it should be noted that this situation may also depend on many variables such as sector, role and regional factors. When we look at hybrid work, we see that the average is less, but I think these people are newcomers because I think Generation Z wants the hybrid work model.

2.Top-Selling Books

Data Source: [Amazon Bestselling Books](#)

Dataset Summary:

The Bestsellers with Categories dataset provides information on top-selling books on Amazon over a span of several years. Each row corresponds to a popular book featured on Amazon's bestsellers list, including details about the book's title, author, user ratings,

price, year, and genre. The dataset allows for analysis of trends in pricing, popularity, and genre preferences over time.

Key Variables:

- **Title (Name):**

Each record includes the name of a bestselling book. Titles range across a wide variety of genres and authors, reflecting diverse reading interests.

- **Author:**

The dataset lists the author(s) of each book, enabling analysis of the most frequent authors or examining how certain authors perform in terms of user ratings and review counts.

- **User Rating:**

A numerical rating that indicates how well-received a book is by readers. Typically on a scale from 1 to 5, these ratings provide insight into reader satisfaction and perceived quality.

- **Reviews:**

The number of customer reviews for each book, reflecting its popularity and the level of engagement it generates. High review counts may indicate broad readership and strong market presence.

- **Price:**

The listed price of the book in U.S. dollars. Price data allows for exploring correlations between price and book popularity, rating, or sales rank.

- **Year:**

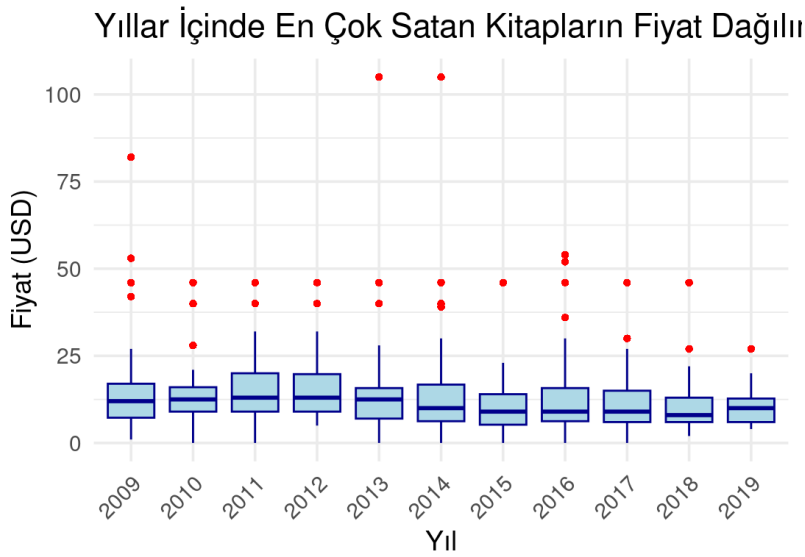
The year the book featured on the bestseller list. This allows for temporal analyses, such as identifying trends in genres, pricing strategies, or shifts in consumer preferences over time.

- **Genre:**

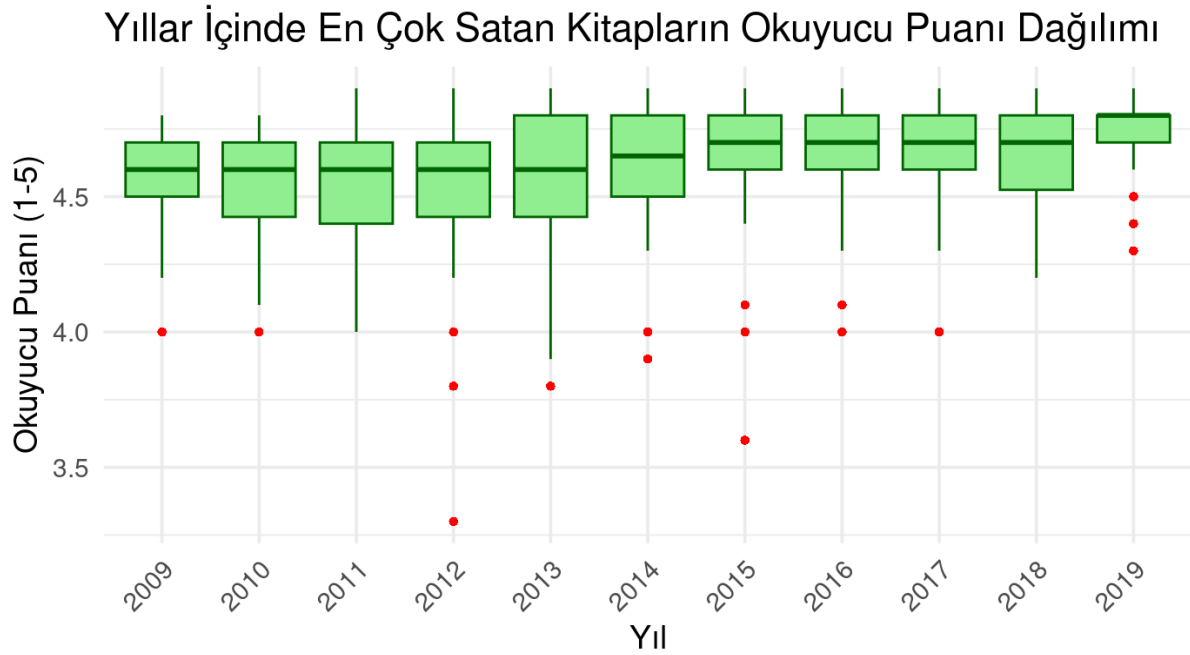
The dataset classifies each book as either Fiction or Non-Fiction. This categorization enables comparison between these two major segments—examining differences in pricing, rating, and popularity, and how these factors evolve year by year.

Questions:

- I. Explore the variation in price distribution of top-selling books over the years.

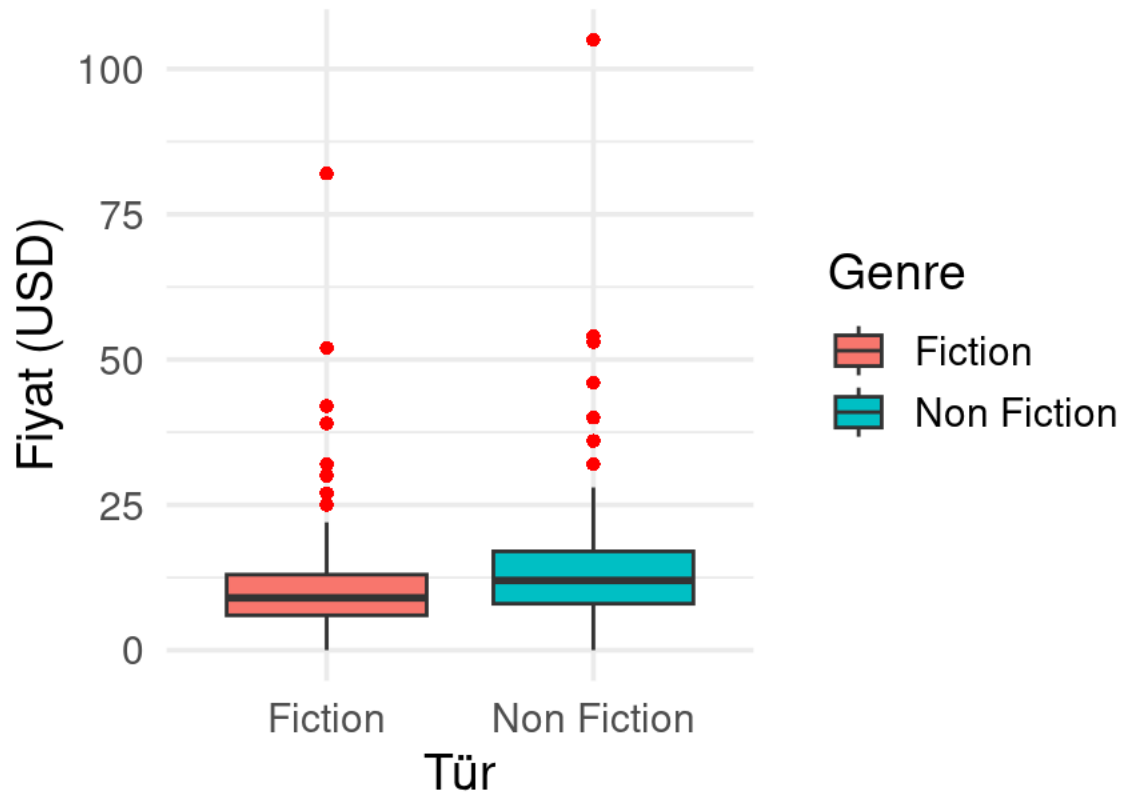


- II. Investigate the change in reader rating distribution of top-selling books over the years.



III. Analyze the price distribution of books by genre.

Fiyat Dağılımı: Türlerine Göre Kitaplar



The chart shows that the best-selling books in both Fiction and Non-Fiction categories are generally clustered in a similar price range, but Non-Fiction has a slightly higher median and a wide price spread. Seeing that very high-priced books occasionally appear in both genres suggests that exceptional pricing strategies may also exist in the market.