

# How Do Stack Overflow Developers Live and Evolve in 2024?

## Introduction

Each year, the Stack Overflow Developer Survey gathers responses from over 90,000 professionals worldwide. Using the CRISP-DM process, we explore trends in compensation, education, and experience to deliver actionable insights for professionals, companies, and educators.

## Data Processing

To ensure the quality of the analysis, missing values in key variables were identified and handled:

Variable	% Missing	Applied Technique
CompTotal	48.4%	Null removal (dropna)
EdLevel	7.1%	Null removal
WorkExp	54.7%	Null removal

Justification: Given the high proportion of missing values in CompTotal and WorkExp, incomplete records were removed to preserve analysis quality. EdLevel had few missing values and was also filtered.

## Key Questions

1. Annual compensation distribution: How do developers' incomes vary?
2. Educational impact: What role do different credentials play in compensation?
3. Career trajectory: How is accumulated experience distributed?

## Key Findings

### 1. Annual Compensation Distribution

- Median: \$110,000 USD
- Q1 (25%): \$60,000 USD
- Q3 (75%): \$250,000 USD

Deep Insight: The wide spread—from \$60,000 in the first quartile to over \$250,000 in the third quartile—reflects a diverse ecosystem, where senior roles and high-cost locations command the highest salaries.

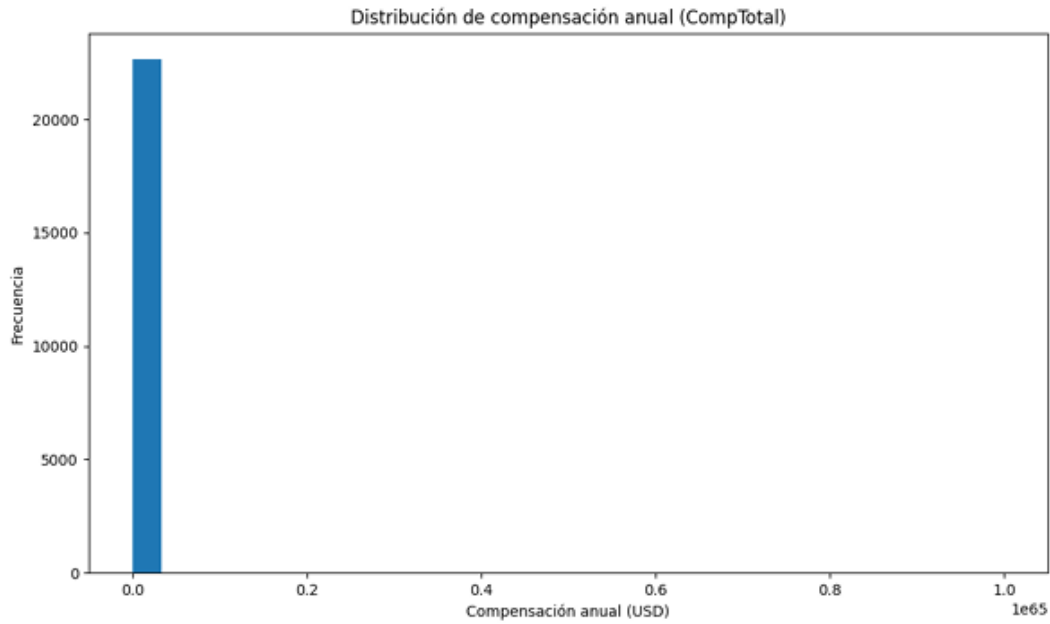


Figure 1. Histogram of CompTotal (30 bins): wide variance highlights market heterogeneity.

## 2. Compensation by Education Level

- Bachelor's degree (B.A., B.S., B.Eng., etc.): \$100,000 USD (median)
- Professional degree (JD, MD, Ph.D., Ed.D., etc.): \$90,000 USD (−10%)
- Master's degree (M.A., M.S., M.Eng., MBA, etc.): \$80,000 USD (−20%)

Deep Insight: Contrary to expectations, bachelor's degree holders show higher median earnings. This may reflect a strong demand for profiles with hands-on experience and practical technical skills.

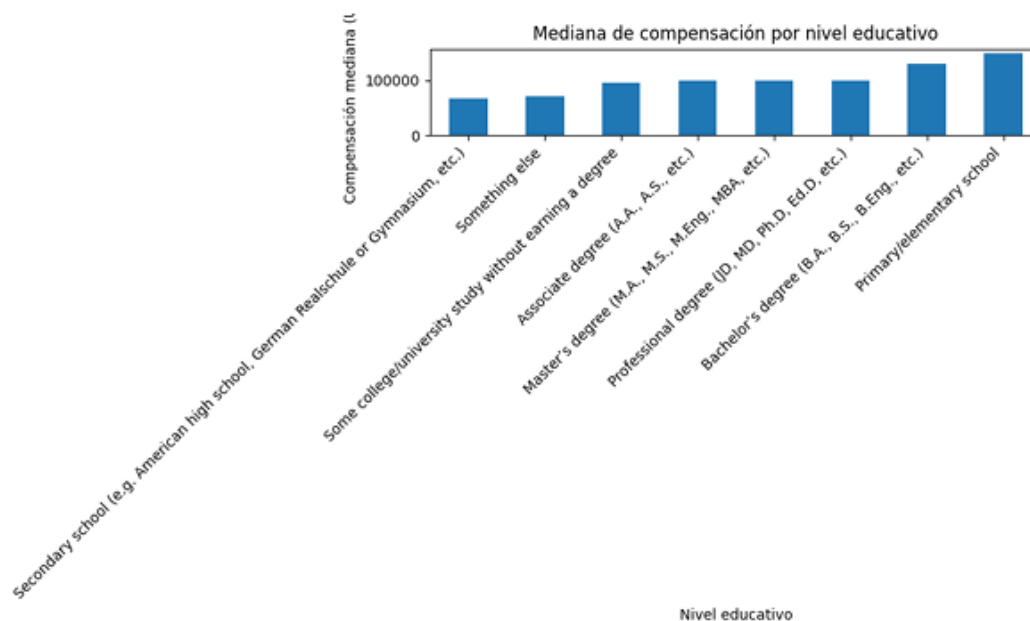


Figure 2. Median CompTotal by EdLevel: ROI across different educational paths.

### 3. Professional Trajectory (Years of Experience)

- Q1 (25%): 5 years
- Median: 10 years
- Q3 (75%): 16 years

Deep Insight: 75% of respondents have more than 16 years of coding experience, indicating a sector with high retention and veteran professionals providing mentorship and stability.

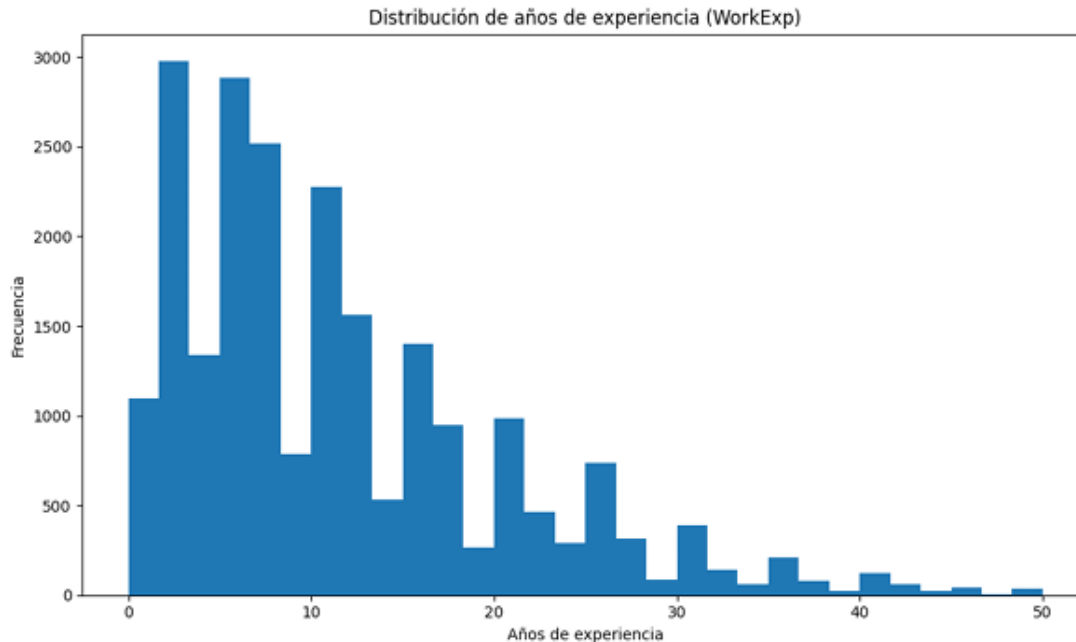


Figure 3. Histogram of WorkExp (30 bins): most respondents have consolidated experience.

### Modularized Code in src/utils.py

To ensure maintainability and follow DRY principles, we grouped functions into a documented module:

```
import pandas as pd
```

```
def cargar_datos(ruta_csv: str) -> pd.DataFrame:
```

```
    """
```

```
    Carga el CSV de Stack Overflow y limpia duplicados.
```

```
    - Elimina duplicados.
```

```
    - Filtra nulos en CompTotal, EdLevel y WorkExp.
```

```
    :param ruta_csv: Ruta al archivo CSV original.
```

```
    :return: DataFrame limpio listo para análisis.
```

```
    """
```

```
    df = pd.read_csv(ruta_csv, low_memory=False)
```

```
    df = df.drop_duplicates()
```

```
    df = df.dropna(subset=['CompTotal', 'EdLevel', 'WorkExp'])
```

```
    return df
```

```
def resumen_estadistico(df: pd.DataFrame, col: str) -> pd.Series:
```

```
    """
```

```
    Obtiene cuartiles y mediana de una columna numérica.
```

```
    :param df: DataFrame de análisis.
```

```
    :param col: Nombre de la columna.
```

```
    :return: Serie con Q1, mediana y Q3.
```

```
    """
```

```
q1 = df[col].quantile(0.25)
med = df[col].median()
q3 = df[col].quantile(0.75)
return pd.Series({'Q1': q1, 'Mediana': med, 'Q3': q3})
```

## Conclusions and Recommendations

1. For Professionals:
  - Focus on hands-on skills and real-world projects.
  - Consider educational pathways alongside practical experience.
2. For Companies:
  - Design career paths with mentoring for senior profiles.
  - Offer continuous technical training.
3. For Academics and Educators:
  - Align curricula with in-demand skills (Cloud, AI, DevOps).
  - Encourage collaborative projects with the industry.

## Best Practices

- Clearly label axes and graph legends.
- Use tables and statistics for direct comparisons.
- Document code with docstrings and follow DRY principles.

## Further Reading

- Data Science Blogs You Need to Check Out: Blog repository with case studies.
- Visualize Data: Advanced visualization techniques.

## Useful Links

- GitHub Repository
  - Tools: pandas, matplotlib, scikit-learn
  - Udacity Data Scientist Course
- <https://github.com/JuanPabloAR-ai/analisis-encuestas-stack-overflow>