Age Distribution Analysis

🌐 **localhost:4746**

Fabricio Lennart

Data Loading

▶ Show Code

```
   age_years  annual_income_usd     gender location_city
0         36           45978.35   Femenino        Madrid
1         29           44882.96  Masculino     Barcelona
2         56           39832.79  Masculino       Sevilla
3         41           40938.62   Femenino      Valencia
4         15           50221.32       Otro        Madrid
```
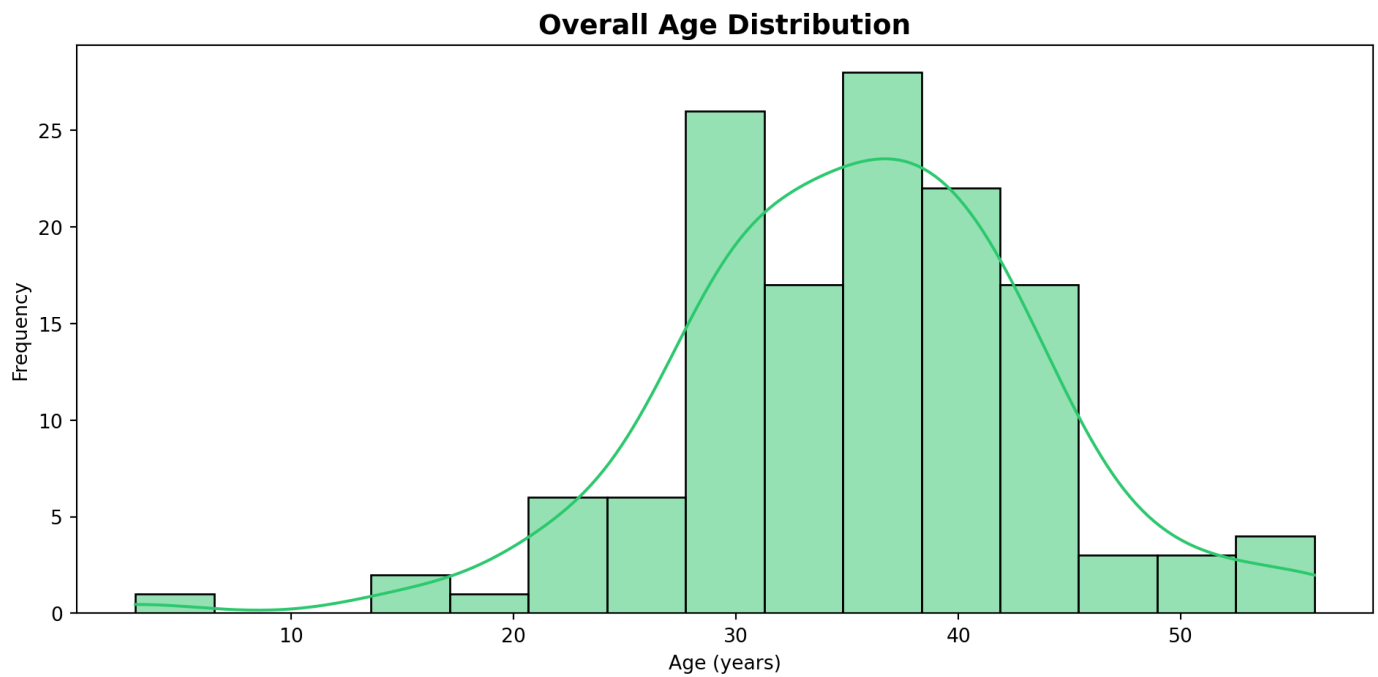
Overall Age Distribution

First, we visualize the **complete dataset** without gender segmentation to understand the general age pattern.

▼ Show Code

```
# Overall distribution
plt.figure(figsize=(10, 5))
sns.histplot(data=data, x="age_years", kde=True, bins=15, color="#2ecc71")
plt.title("Overall Age Distribution", fontsize=14, fontweight='bold')
plt.xlabel("Age (years)")
plt.ylabel("Frequency")
plt.tight_layout()
plt.show()
```
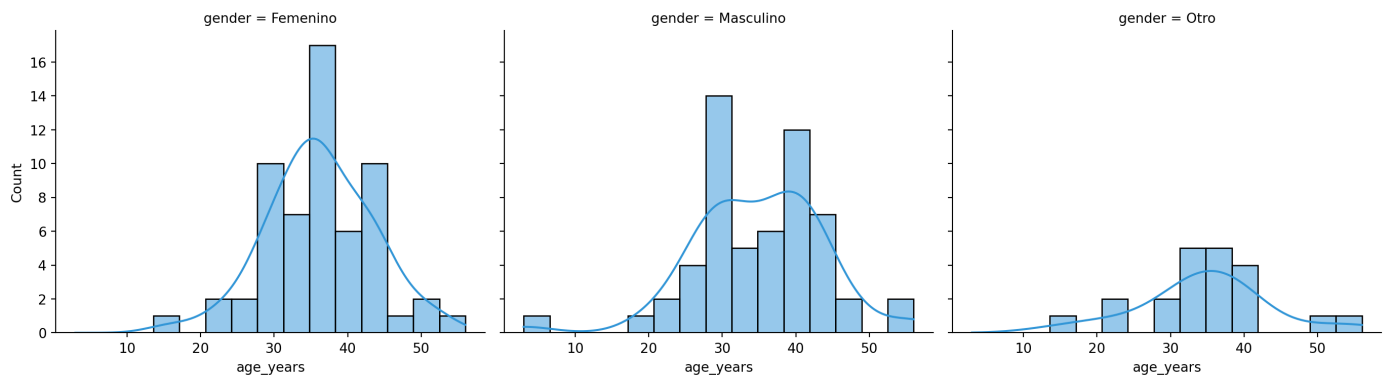
Overall Age Distribution

Age Distribution by Gender

Now we compare distributions across three gender categories using **faceted histograms**.

▼ Show Code

```
# Distribution by gender
sns.displot(
    data=data,
    x="age_years",
    col="gender",
    kde=True,
    bins=15,
    height=4,
    aspect=1.2,
    color="#3498db"
)
plt.tight_layout()
plt.show()
```

Key Findings

Female: Unimodal Gaussian distribution centered at 30-40 years with low variability.

Male: Bimodal pattern with peaks at 30-35 and 40-45 years, suggesting two distinct age cohorts.

Other: Platykurtic distribution with higher dispersion and no clear central peak.

Built with pandas, seaborn & matplotlib