1. Conclusion Explained

The analysis concludes that salary and age are the primary drivers of EV affordability, with high-income individuals aged 35+ most likely to afford EVs. The model accurately identifies potential buyers (F1-score: 0.91 for class "Yes") but struggles with budget-constrained groups (F1-score: 0.57 for class "No"). This imbalance arises because the dataset contains fewer examples of customers who cannot afford EVs (4 vs. 16 in the test set).

Key Insights:

- Salary Threshold: Individuals need a salary ≥ 1.5× the EV price to afford it.
- Age Correlation: Affordability increases with age (peak at 35+), likely due to higher career stability and income.
- Minority-Class Limitations: The model biases toward predicting "Yes" due to imbalanced data.

2. Process: Models, Frameworks & Libraries

Tech Stack

- Language: Python
- Libraries:
  - `pandas`: Data loading/manipulation.
  - `scikit-learn`: Machine learning (encoding, model training, evaluation).
- Model: Random Forest Classifier (ensemble method using decision trees).

Workflow

1. Data Preparation:
   - Created target variable `Can Afford EV` (`Yes`/`No`) based on salary-to-price ratio (`Total Salary ≥ 1.5 * Price`).
   - Encoded categorical features (`Profession`, `Education`, etc.) into numerical values using `LabelEncoder`.
2. Model Training:
   - Split data into training (80%) and testing sets (20%).
   - Trained a `RandomForestClassifier` to predict affordability.

3. Evaluation:
    - Generated a `classification_report` with precision, recall, and F1-scores.
    - Analyzed feature importances to identify key predictors.
4. Inference:
    - Tested synthetic age-based profiles to simulate affordability scenarios.

3. Graphs, Visualizations & Implications

While the provided code lacks visualizations, key outputs imply:

A. Classification Report

```
        precision    recall  f1-score   support
0       0.67      0.50      0.57        4  → "Cannot Afford"
1       0.88      0.94      0.91       16  → "Can Afford"
```

- Implication: High precision/recall for class `1` (afford) but low recall for class `0` (cannot afford). The model misses 50% of budget-constrained users.

B. Feature Importances

Feature	**Importance**
Age	30.5%
Total Salary	30.3%
No of Dependents	16.9%
Education	7.2%
Marrital Status	6.9%
Profession	4.1%
Personal loan	4.1%

- Implication:
  - Age and salary dominate predictions (60.8% combined).
  - Dependents reduce affordability (e.g., childcare costs).
  - Profession/loan status are minor factors.

C. Synthetic Age-Group Predictions

Age	**Prediction**
25	Cannot Afford EV
35	Can Afford EV
45	Can Afford EV
55	Can Afford EV

- Implication: Affordability spikes after age 35, aligning with career progression.

4. Data-Driven Solutions for the Company

1. Target High-Probability Groups:
   - Focus marketing on salaried professionals aged 35–60 with salaries ≥ 2M.
   - Offer loyalty discounts to customers in this segment.
2. Address Younger Demographics:
   - Launch financing schemes (EMI, subsidies) for customers aged 21–34.
   - Partner with banks for low-interest EV loans.
3. Improve Model Reliability:
   - Collect more data for budget-constrained users to rebalance classes.
   - Add features: existing debts, location (urban/rural), fuel costs.
4. Operational Adjustments:
   - Use the model to pre-qualify leads in sales pipelines.

- Develop dynamic pricing: Adjust EV packages based on predicted affordability.

5. GitHub Repository

Access the full code, dataset, and detailed analysis here:

🔗 EV Affordability Prediction GitHub Link