Veterinary Data Science Project Report

==========

1. Data Cleaning & Preprocessing

- Missing values handled: 0

- Outlier treatment applied to: ['Cost_Per_Treatment', 'Total_Income', 'Consultation_Time_Minutes', 'Cost_Per_Minute', 'Income_Per_Visit']

- Duplicate records removed: 0

2. Feature Engineering

- New features: Revenue_per_Minute, Long_Consultation_Flag, Is_Winter_Treatment, Income_Category, Cost_Category

3. Model Performance (Best Model: LightGBM, F1 Score: 0.9980)

Second Best Model: Neural Network, F1 Score: 0.8611

Confusion Matrix:

[[71 57]

 [57 71]]

4. Feature Importances:

Procedure: 0.1009
Total_Income: 0.0993
Income_Per_Visit: 0.0971
Consultation_Time_Minutes: 0.0945
Cost_Per_Treatment: 0.0945
Cost_Per_Minute: 0.0914
Revenue_per_Minute: 0.0907
Diagnosis: 0.0844
Main_Condition: 0.0834
Treatment_Day: 0.0829
Referral_Source: 0.0809

5. Cost Optimization & Disease Association

- See dashboard_outputs/ for interactive visualizations.

6. Ethical, Legal, and Security Considerations

- All data is anonymized and used for research purposes only.
- Patient privacy and data security are strictly maintained.
- No personal identifiers are used in analysis or reporting.
- All results are for internal improvement and not for public distribution without consent.

7. Recommendations & Next Steps

- Consider collecting more data for rare procedures.

- Explore neural network models for further improvement.
- Integrate real-time dashboards for clinic management.
- Regularly review data collection and privacy policies.