Case Study: A Data-Driven Analysis of a Fitness Application

MSE803: Data Analytics - Assessment 2

Supervisor: Dr. Mukesh Mishra
Prepared by Thanushka Wickramrachchi
Student ID: 270599091
Jul 13, 2025

YOOBEE COLLEGE OF CREATIVE INNOVATION

Introduction & Objective

- **Analyze user behavior in a fitness app**

  How users behave, how much they move, and how the app can support them better

- **Predict calories burned and activity levels**

  Built models to predict how many calories someone burns

- **Group users into segments (clustering)**

  grouped users based on their behavior using clustering.

- **Suggest smart app improvements**
- **Focus on ethical and culturally respectful design**

  Focused on ensuring the app respects people’s privacy and culture.

Dataset Overview

- **5,000 User Records**
  - dataset includes information from 5,000 people who use a fitness tracking app
- **8 Key Attributes**
  - Each person has 8 types of data
  - Age, Gender, Location, Activity Level, App Sessions, Distance, Calories
- **0 Missing Values**
  - Data is already clean
  - No need to fix or fill anything before starting the analysis

Data Preprocessing

- **Preparing the Data**
  - Before using the data for machine learning, had to prepare it
- **Encoded text data into numbers**
  - Location, gender, and activity level **-** Changed text data into numbers using encoding
- **Grouped age into three ranges:** 18–30, 31–45, 46–60
  - grouped ages into three categories - This helps. to discover trends or differences between different age groups
- **Standardized numbers**
  - Important for fair results in clustering and modeling.

Exploratory Data Analysis

- **More sessions = more distance & calories**
  - Explored the data to understand general patterns
  - key findings was that people who use the app more often usually walk more and burn more calories.
- **Active users burn more calories**
  - Active users - burned more calories than those who are sedentary
- **Age doesn’t affect calorie burn much**
  - That means young and older users can be equally active
- **Data is well balanced (gender, location, activity)**
  - dataset was balanced. There were equal numbers of male and female users and a good mix from urban, suburban, and rural areas

Exploratory Data Analysis (CONT.)

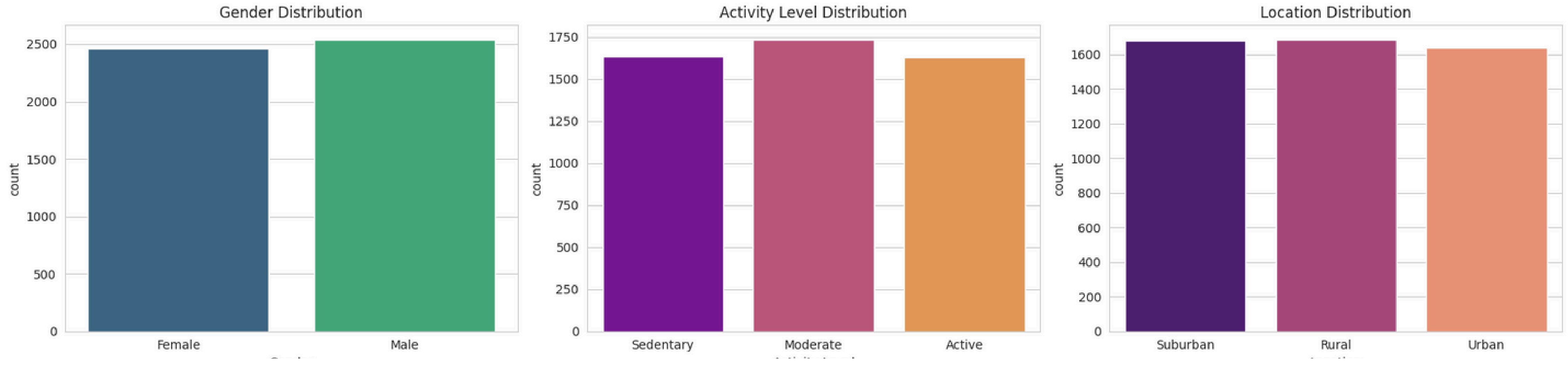

Figure 1 : Distribution of categorical features — Gender, Activity Level, and Location

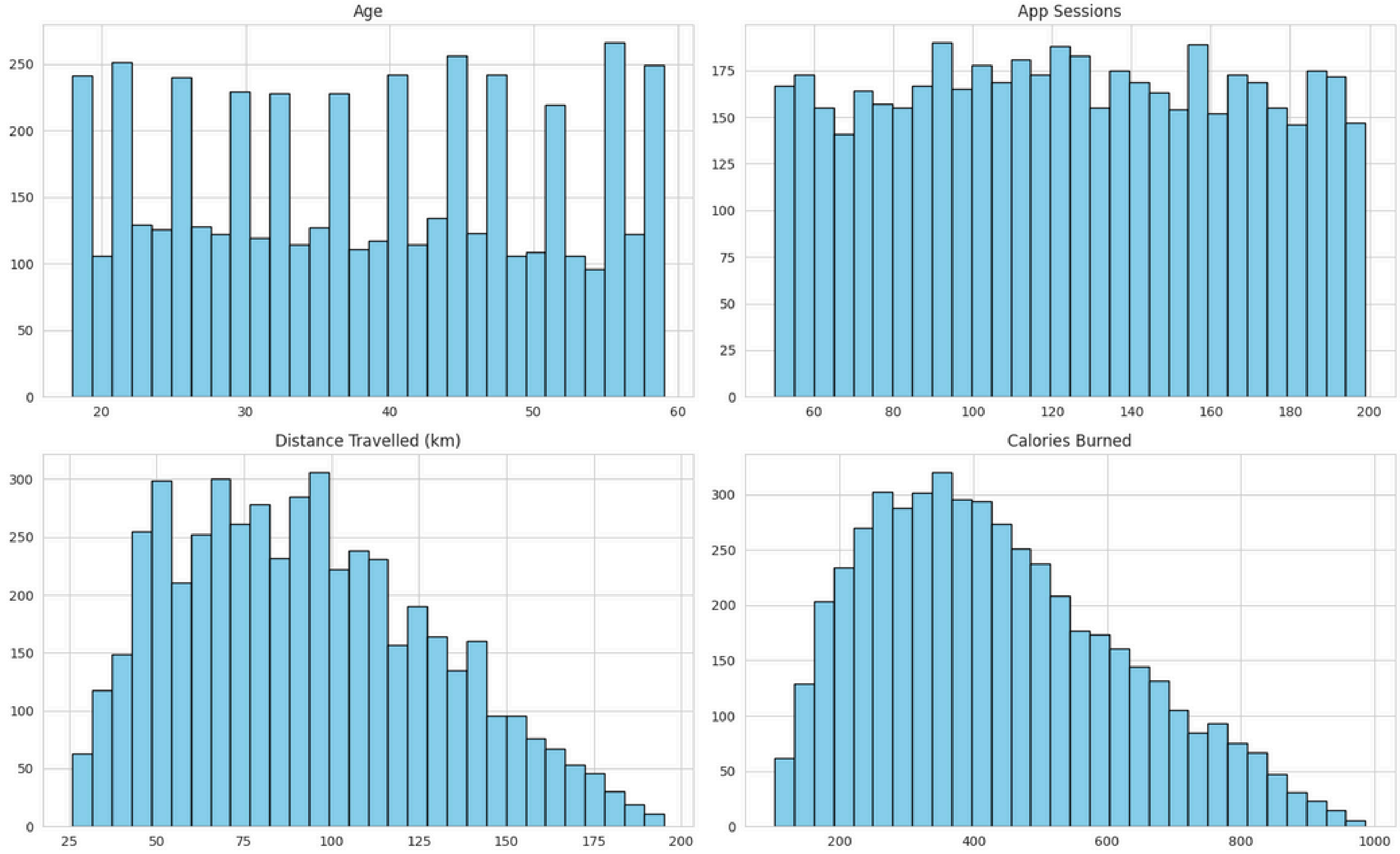

Figure 2: Histograms showing the distribution of Age, App Sessions, Distance Travelled (km), and Calories Burned

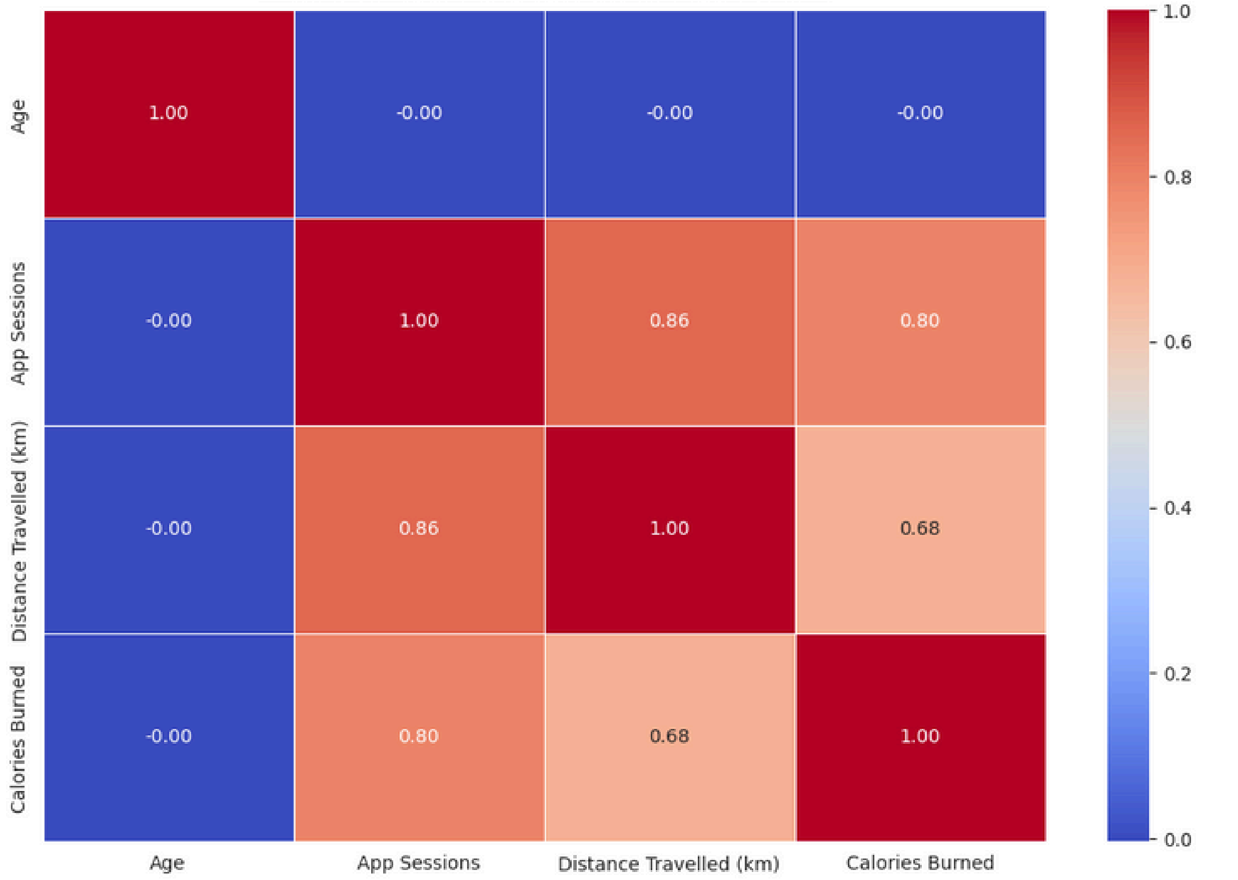

Figure 3 : Correlation matrix showing linear relationships among Age, App Sessions, Distance Travelled, and Calories Burned

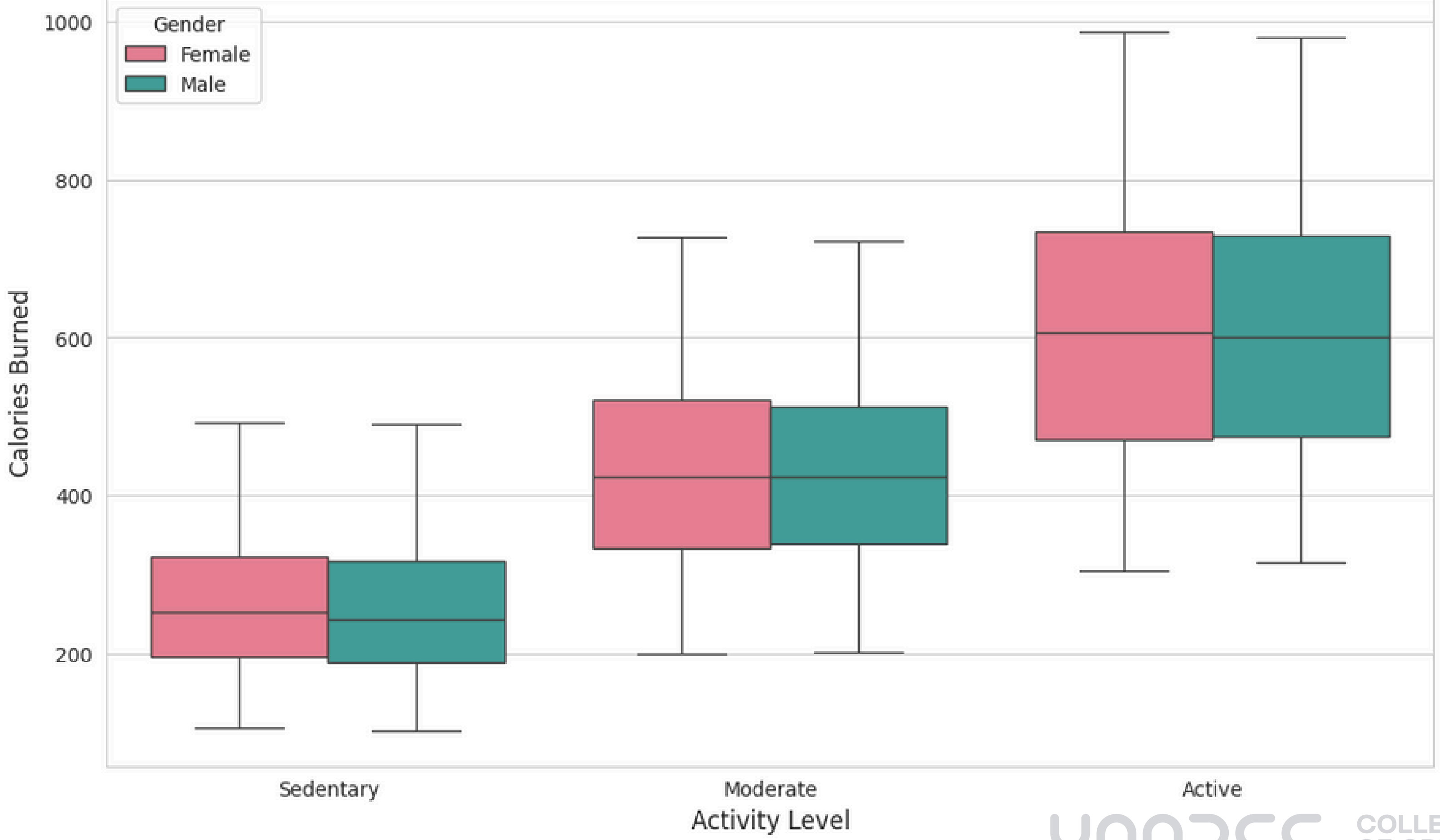

Figure 4: Calories Burned by Activity Level and Gender (Boxplot)

Predictive Modeling

- **Used two different models**
  - Predict how many calories a user will burn
- **Linear Regression (basic)**
  - First, I tried this - simple and easy to understand, but not very powerful for complex data
- **Random Forest (advanced)**
  - Then, I used Random Forest. Usually this model is more powerful and works better

Model	Accuracy (R^2 Score)	Prediction Error (MAE)
Linear Regression (Simple)	**0.64**	**93.2 Calories**
Random Forest (Complex)	0.6	96.1 Calories

Table 1: Model Performance Comparison: Linear Regression vs. Random Forest

- In this case, both models gave good results; the simple model did slightly better.
- I thought a complex model would work best — but it was wrong. The simple linear regression model was actually more accurate.
- This provides a strong foundation for building **predictive features**.

User Segmentation with Clustering

- **Used K-Means (4 clusters)**
  - I used a technique called K-Means Clustering to group users based on their real behavior
- **Groups found:**
  - Cluster 1 (Yellow): Young Power Users (very active)
  - Cluster 3 (Purple): Older Power Users (highly active but older)
  - Cluster 0 (Blue): Low-Engagement Active Users (Steady, but room to grow)
  - Cluster 2 (Green): Older Casual Users (Lowest frequency within the active base)

- These user groups help the app team.
- Instead of showing the same thing to everyone, the app can be personalized.
- E.g., Inactive users can get motivational messages, and active users can see advanced features.

User Segmentation with Clustering (CONT.)

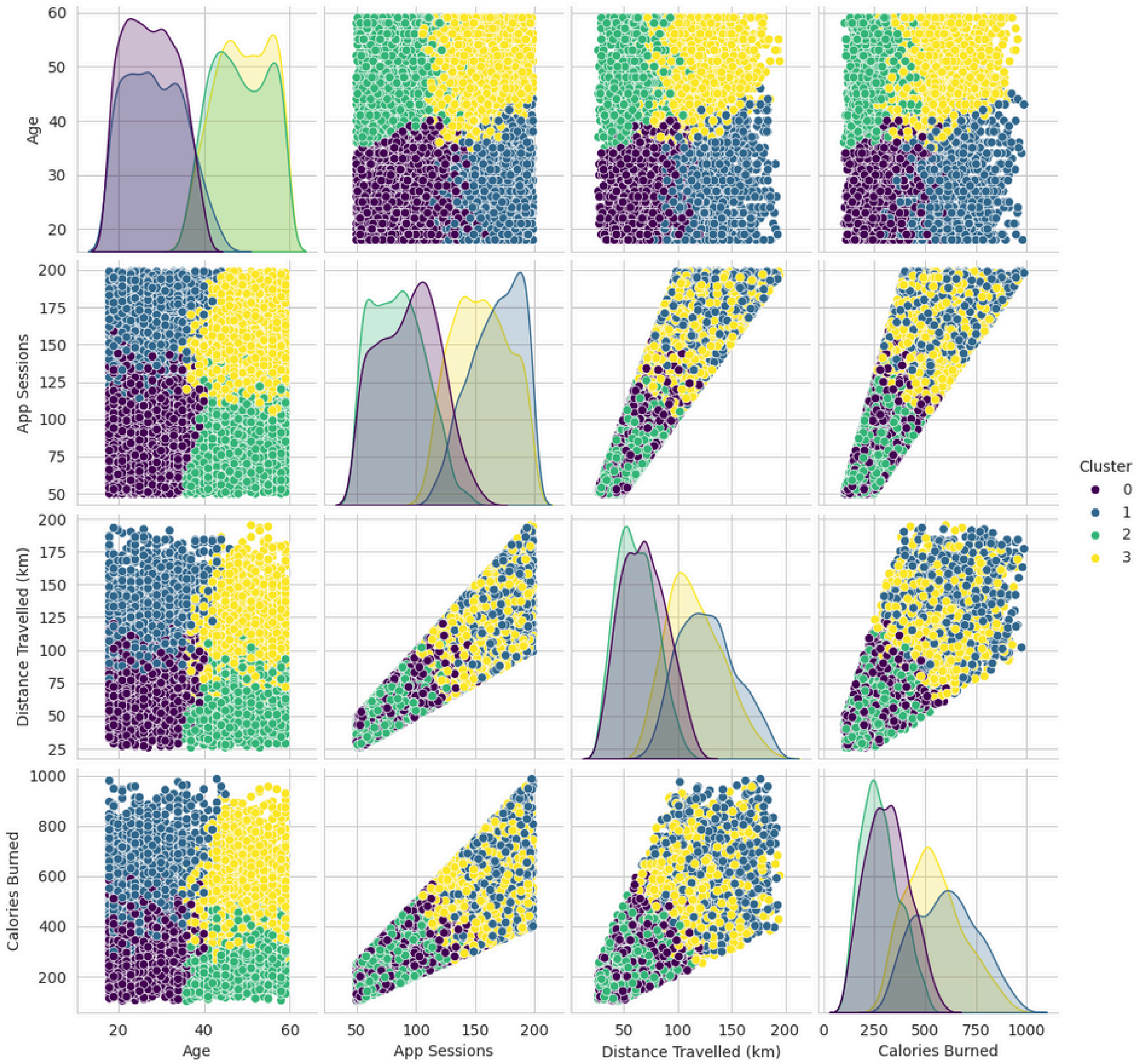

Figure 5: Pair Plot of User Segments Colored by Cluster

Engineering Implications

- **Two Smart Features We Can Build:**

1. **Goal Prediction -** Build real-time calorie prediction feature
    - **Predict how many calories a user will burn**
        - We can show users how many calories they'll burn before they start a workout (Eg - You'll burn 550 calories for 10km)
    - **Based on: Simple Linear Regression model**
    - **Benefit: Fast, accurate, and easy to use**
2. **Adaptive User Interface (UI) -** Show different UI for each user group
    - **App screen changes based on user type**
        - The app screen can change based on the user's group
    - **Based on: 4 user groups from clustering**
    - **Benefit: Easy for beginners, advanced for active users**
        - giving more tips to new users and advanced stats to active ones.

Engineering Implications (CONT.)

- **Use A/B testing based on user type**
  - we can test if motivational notifications work.
  - Older Casuals - send a notification to half of them and see if their engagement increases.
  - This is a much smarter way to improve the app

- **Create a live dashboard for developers**
  - developers and managers can use a live dashboard to track user behavior, engagement, and the **effects of new features**.

Ethical & Cultural Considerations

- **Respect user privacy**
  - Tell users what data is collected and why, and keep their info safe.
- **Avoid bias in predictions**
  - the app should not assume older users are always inactive or make unfair recommendations.
- **Make features culturally sensitive**
  - Some users may not want competitive messages or may prefer different styles of motivation based on age or culture.
- **Give users control over their data**
  - give users control — let them see, change, or delete their data.
- **Follow standards like GDPR**
  - to keep everything legal and fair.

Strategic Recommendations & Conclusion

- **Build real-time calorie prediction**
  - Add calorie prediction using the model we built — it's accurate and useful.
- **Use cluster-based personalized UI**
  - Give advanced features to power users and simpler screens to inactive ones.
- **Run A/B tests by user type**
  - Run A/B tests by user group to find out what features or messages work best.
- **Launch internal BI dashboard**
  - Build a business intelligence (BI) dashboard for the team — so they can track users and features in real time.
    - http://localhost:8501/
- **Follow ethical data practices**
  - most importantly, protect user privacy and be fair with data — follow rules and stay transparent.
- By using smart data analysis and respecting user privacy, we can build a better app that is smart, safe, and helpful for everyone.

References

- Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. Proceedings of the 1st Conference on Fairness, Accountability and Transparency, 81, 77–91. https://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf
- Cavoukian, A. (2009). Privacy by design: The 7 foundational principles [Technical report]. Information and Privacy Commissioner of Ontario. https://student.cs.uwaterloo.ca/~cs492/papers/7foundationalprinciples_longer.pdf
- Disterer, G. (2013). ISO/IEC 27000, 27001 and 27002 for information security management. Journal of Information Security, 4(2), 92–100. https://doi.org/10.4236/jis.2013.42011
- European Union. (2016). General Data Protection Regulation [Official legal text]. https://eur-lex.europa.eu/eli/reg/2016/679/oj
- Friedman, B., & Nissenbaum, H. (1996). Bias in computer systems. ACM Transactions on Information Systems (TOIS), 14(3), 330–347. https://doi.org/10.1145/230538.230561
- Papageorgiou, A., Strigkos, M., Politou, E., Alepis, E., Solanas, A., & Patsakis, C. (2018). Security and privacy analysis of mobile health applications: The alarming state of practice. IEEE Access, 6, 9390–9403. https://doi.org/10.1109/ACCESS.2018.2799522
- U.S. Department of Health & Human Services. (2013). Summary of the HIPAA privacy rule. https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html

Thank You

YOOBEE COLLEGE OF CREATIVE INNOVATION