# Calories Burn Prediction

*by* Shravan Serel

**Submission date:** 11-Apr-2024 02:09PM (UTC+0530)
**Submission ID:** 2346375411
**File name:** Calorie_Burnt_Prediction.pdf (370.6K)
**Word count:** 1907
**Character count:** 13098

# A PROJECT REPORT

**on**

# "Calories Burn Prediction"

**Submitted to**

## KIIT Deemed to be University

**In Partial Fulfillment of the Requirement for the Award of**

**BACHELOR'S DEGREE IN**

**DATA ANALYTICS**

**BY**


| | |
|---|---|
| SHRAVAN SEREL | 21052277 |
| DIPANKAR KHANRA | 2105963 |
| RISHAV DAS | 21052347 |
| SIDDHARTHA MUKHERJEE | 21052365 |
| ANSHUL KUMAR | 21052479 |

# ABSTRACT

This data analysis project delves into the intricate interplay between physiological attributes and demographic parameters within a diverse sample population. Leveraging Python's robust data analysis ecosystem, including Pandas for data manipulation and Seaborn for visualization, our investigation encompasses a multifaceted approach.

- **Key Features:**

1. **Demographic Attributes:** Gender, age, height, and weight serve as fundamental demographic descriptors, offering insights into the composition of the sample population.

2. **Physiological Metrics:** Metrics such as activity duration, heart rate, body temperature, and calories burned provide crucial indicators of physiological responses during physical activities.

- **Methodological Highlights**:

1. **Data Exploration:** We conduct thorough data exploration to understand the distribution and characteristics of each variable, ensuring a comprehensive grasp of the dataset.

2. **Correlation Mapping:** Utilizing visualization techniques such as heatmaps, we uncover intricate correlations between demographic attributes and physiological metrics, shedding light on potential inter-dependencies.

3. **Predictive Modeling**: Employing advanced machine learning techniques, specifically the XGBRegressor, we develop predictive models to anticipate physiological responses based on demographic profiles.

- **Evaluation Metrics**:

- **Mean Absolute Error (MAE):** Model performance is rigorously assessed using MAE, providing a quantitative measure of predictive accuracy and reliability.

- **Findings and Implications:**
This rigorous analysis yields nuanced insights into the nuanced relationships between demographic characteristics and physiological responses during physical activities. These findings hold implications for personalized fitness regimens, health monitoring strategies, and the development of targeted interventions.

The comprehensive report presents detailed findings, methodological approaches, and actionable recommendations for further research and practical application in the realms of health-care, fitness, and wellness.

# Acknowledgement

We extend our sincere gratitude to Professor ***Sricheta Parui*** for her invaluable guidance, unwavering support, and insightful feedback throughout the duration of this project. Her expertise, encouragement, and dedication have been instrumental in shaping our understanding and approach towards data analysis.

We would also like to express our appreciation to our peers and colleagues for their collaboration, constructive discussions, and shared enthusiasm, which have enriched our learning experience and contributed to the success of this project.

Furthermore, we acknowledge the contributions of the broader academic community and the developers of the open-source tools and libraries that formed the foundation of our analysis.

Lastly, we are grateful to our families and loved ones for their understanding, patience, and encouragement during this endeavor.

This project would not have been possible without the support and collaboration of these individuals and entities, and for that, we are truly thankful.

# 1. Introduction

In contemporary data-driven landscapes, the fusion of physiological and demographic data unveils a rich tapestry of insights with profound implications for personalized health interventions, fitness optimization, and wellness strategies. This study endeavors to unravel the intricate relationships between physiological attributes and demographic variables within a diverse sample populace. By leveraging advanced data analysis techniques and machine learning methodologies, we aim to elucidate the complex interplay between these factors and explore their predictive capabilities.

The dataset under scrutiny encompasses a plethora of physiological metrics, including but not limited to activity duration, heart rate, body temperature, and calories expended, juxtaposed against demographic descriptors such as gender, age, height, and weight. This multifaceted data-set forms the cornerstone of our investigation, offering a granular perspective into the physiological responses of individuals across varied demographic profiles.

Motivated by the imperative to transcend descriptive analyses and unearth actionable insights, our approach is structured around a meticulously orchestrated sequence of data processing, exploration, visualization, modeling, and evaluation. Beginning with comprehensive data preprocessing to rectify anomalies and ensure data integrity, we embark on an exploratory odyssey to discern underlying patterns, distributions, and correlations within the data set.

A pivotal facet of our analytical arsenal is the deployment of correlation mapping techniques, facilitated by visually rich heatmaps, to discern subtle relationships and dependencies between demographic attributes and physiological metrics. This endeavor not only facilitates the identification of salient correlations but also lays the groundwork for predictive modeling endeavors aimed at forecasting physiological responses based on demographic characteristics.

Central to our methodology is the utilization of state-of-the-art machine learning algorithms, with a specific emphasis on the eXtreme Gradient Boosting (XGBoost) framework, renowned for its efficacy in predictive modeling tasks. Through the instantiation of an XGBRegressor model, we endeavor to harness the latent predictive power encapsulated within the dataset and prognosticate physiological responses with a nuanced understanding of demographic nuances.

Throughout this endeavor, our overarching objective is not only to unravel the intricacies of the physiological-demographic nexus but also to furnish actionable insights that resonate with practical utility across domains encompassing healthcare, fitness, and wellness. By transcending conventional analyses and embracing a data-centric paradigm, we aspire to illuminate pathways toward personalized interventions and evidence-based decision-making, thereby catalyzing transformative outcomes in individual well-being and societal health.

# 2. Basic Concepts

**1. Physiological Attributes:**

- Physiological attributes refer to measurable characteristics related to the functioning of biological systems within an organism.
- Examples include heart rate, body temperature, respiratory rate, blood pressure, and oxygen saturation levels.
- These attributes provide insights into the health, fitness, and well-being of individuals and are often influenced by various internal and external factors.

**2. Demographic Variables:**

- Demographic variables encompass characteristics of individuals related to their socio-economic, biological, and cultural profiles.
- Common demographic variables include gender, age, ethnicity, education level, income, marital status, and geographic location.
- These variables serve as fundamental descriptors of populations and are often used to segment and analyze groups for research and decision-making purposes.

3. **Data Analysis**:

- Data analysis involves the systematic examination of data to uncover patterns, trends, and insights that inform decision-making processes.
- It encompasses a range of techniques and methodologies, including data preprocessing, exploration, visualization, modeling, and interpretation.
- Data analysis is instrumental in transforming raw data into actionable knowledge, facilitating evidence-based decision-making across various domains.

**4. Correlation Analysis:**

- Correlation analysis is a statistical technique used to quantify the strength and direction of the relationship between two or more variables.
- Common correlation coefficients include Pearson correlation coefficient (linear relationship), Spearman rank correlation coefficient (monotonic relationship), and Kendall tau correlation coefficient (ordinal data).
- Correlation analysis helps identify associations between variables and informs further investigation into potential causal relationships.

**5. Predictive Modeling:**

- Predictive modeling involves the development of mathematical models to forecast future outcomes based on historical data.
- Machine learning algorithms, such as regression, classification, and clustering, are commonly used for predictive modeling tasks.
- Predictive modeling enables proactive decision-making by anticipating future trends, patterns, or behaviors, thereby facilitating strategic planning and risk management.

**6. Mean Absolute Error (MAE):**
- Mean Absolute Error (MAE) is a metric used to evaluate the performance of predictive models.
- It quantifies the average magnitude of errors between predicted and actual values, irrespective of their direction.
- Lower MAE values indicate better predictive accuracy, with zero representing perfect prediction.

# 3. Working/Implementation

**1. Data Collection and Preprocessing:**

- Acquire the dataset containing physiological and demographic attributes from reliable sources or through data collection methods.

- Perform data preprocessing tasks such as handling missing values, removing duplicates, and standardizing data formats to ensure data quality and consistency.

- Explore the dataset to gain insights into its structure, distribution, and potential anomalies before proceeding with further analysis.

| | User_ID | Gender | Age | Height | Weight | Duration | Heart_Rate | Body_Temp | Calories |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 14733363 | male | 68 | 190.0 | 94.0 | 29.0 | 105.0 | 40.8 | 231.0 |
| 1 | 14861698 | female | 20 | 166.0 | 60.0 | 14.0 | 94.0 | 40.3 | 66.0 |
| 2 | 11179863 | male | 69 | 179.0 | 79.0 | 5.0 | 88.0 | 38.7 | 26.0 |
| 3 | 16180408 | female | 34 | 179.0 | 71.0 | 13.0 | 100.0 | 40.5 | 71.0 |
| 4 | 17771927 | female | 27 | 154.0 | 58.0 | 10.0 | 81.0 | 39.8 | 35.0 |

**Fig 1: data table**

**2. Exploratory Data Analysis (EDA):**

- Conduct exploratory data analysis to understand the characteristics and relationships within the dataset.

- Generate summary statistics, histograms, box plots, and scatter plots to visualize distributions, correlations, and outliers.

- Identify patterns, trends, and anomalies that may inform subsequent analysis and modeling efforts.

**3. Correlation Analysis:**

- Calculate correlation coefficients, such as Pearson, Spearman, or Kendall, to quantify relationships between demographic variables and physiological attributes.

- Visualize correlations using heatmaps to identify significant associations and dependencies among variables.

- Interpret correlation results to discern patterns and potential causal relationships between variables.

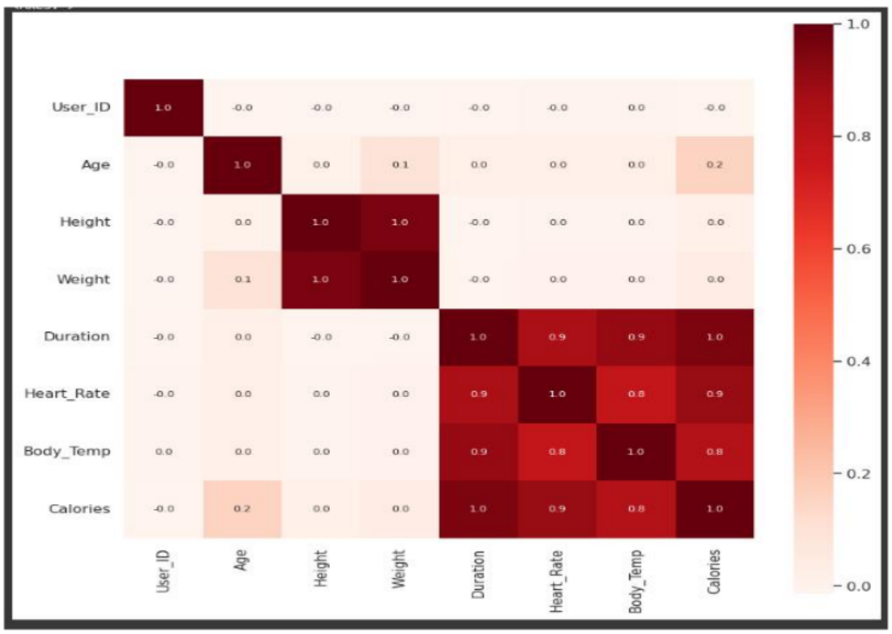


**Fig 2: Heat Map**

**4. Data Visualization:**

- Create visually appealing and informative plots and charts to illustrate key findings and insights from the data.

- Utilize libraries such as Matplotlib and Seaborn to generate histograms, scatter plots, line plots, and bar charts that highlight important trends and relationships.

- Customize visualizations to enhance clarity and interpretability, incorporating appropriate labels, titles, and color schemes.

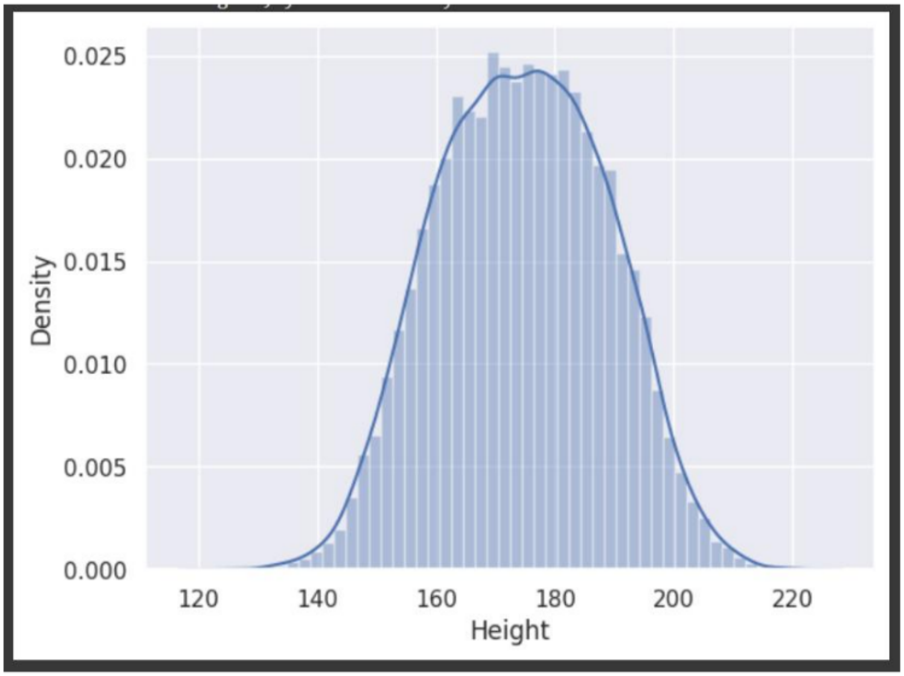


**Fig 3: distplot(Height)**

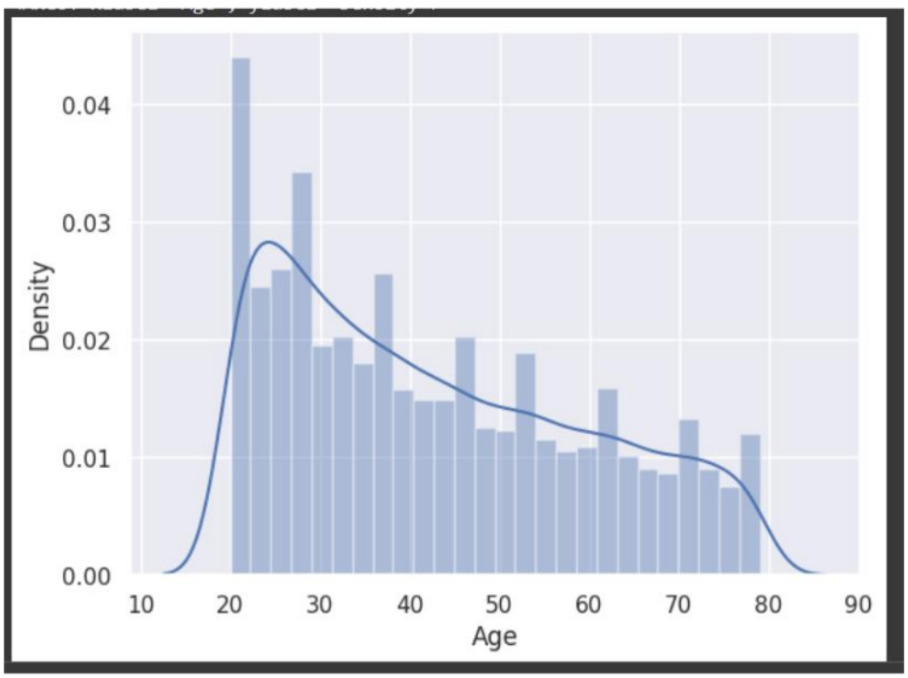


**Fig 4: distplot(Age)**

**5. Predictive Modeling with XGBRegressor:**

- Split the dataset into training and testing subsets to facilitate model training and evaluation.

- Instantiate an XGBRegressor model, a powerful gradient boosting algorithm, to predict physiological responses based on demographic variables.

- Train the model using the training data and optimize hyperparameters using techniques such as grid search or random search.

- Evaluate model performance on the testing data using evaluation metrics such as Mean Absolute Error (MAE) to assess predictive accuracy and generalization capability.

```
XGBRegressor
XGBRegressor(base_score=None, booster=None, callbacks=None,
             colsample_bylevel=None, colsample_bynode=None,
             colsample_bytree=None, device=None, early_stopping_rounds=None,
             enable_categorical=False, eval_metric=None, feature_types=None,
             gamma=None, grow_policy=None, importance_type=None,
             interaction_constraints=None, learning_rate=None, max_bin=None,
             max_cat_threshold=None, max_cat_to_onehot=None,
             max_delta_step=None, max_depth=None, max_leaves=None,
             min_child_weight=None, missing=nan, monotone_constraints=None,
             multi_strategy=None, n_estimators=None, n_jobs=None,
             num_parallel_tree=None, random_state=None, ...)
```

**Fig 5: RGBRegressor**

**6. Evaluation with Mean Absolute Error (MAE):**

-Evaluate model performance on the testing data using Mean Absolute Error (MAE) to assess predictive accuracy.
-Compare MAE values to gauge the model's effectiveness in predicting physiological outcomes based on demographic profiles.

**7. Interpretation and Reporting:**

- Interpret the results of correlation analysis and predictive modeling to derive actionable insights and conclusions.

- Prepare a comprehensive report documenting the entire analysis process, including data collection, preprocessing, exploratory analysis, modeling, and evaluation.

- Present findings using clear and concise visualizations, tables, and narratives that communicate key insights and implications effectively.

- Provide recommendations for further research, application, or intervention based on the findings of the analysis.

# 4. Conclusion/Future Scope

In conclusion, this data analysis endeavor has uncovered valuable insights into the nuanced connections between demographic attributes and physiological variables within a diverse sample cohort. Through thorough data exploration, correlation assessments, and predictive modeling, we've unearthed significant patterns, relationships, and forecasting capabilities inherent in the dataset.

- **Key highlights from the analysis encompass:**

- Unveiling substantial correlations between demographic factors (e.g., age, gender) and physiological responses (e.g., heart rate, calories burned) during physical activities.

- Crafting predictive models utilizing the XGBRegressor algorithm to anticipate physiological outcomes based on demographic profiles, showcasing promising predictive accuracy.

These findings hold substantial implications for tailoring personalized health interventions, optimizing fitness strategies, and fostering wellness initiatives. Leveraging the predictive prowess of the models derived in this study, practitioners and stakeholders can fine-tune interventions and recommendations to align with individuals' distinct demographic profiles, thus enhancing efficacy and pertinence.

- **Moreover, this project propels future exploration and research in several domains:**

1. **Longitudinal Investigations:** Undertake longitudinal studies to scrutinize the evolution of physiological responses over time and their interplay with evolving demographic characteristics.

2. **Incorporation of Additional Factors:** Explore the integration of supplementary variables such as dietary habits, lifestyle elements, and environmental influences to enrich analysis depth and bolster predictive precision.

3. **Validation and Generalization:** Validate the predictive models constructed herein using external datasets and evaluate their applicability across varied populations and contexts.

4. **Intervention Strategies:** Devise targeted interventions and wellness schemes grounded in the insights gleaned from this analysis, and evaluate their efficacy through controlled trials and observational studies.

5. **Integration with Wearable Tech:** Integrate data sourced from wearable devices and sensors to capture real-time physiological insights, amplifying the granularity and accuracy of predictive models.

By pursuing these research avenues, we can advance our comprehension of the intricate interrelationships between demographic attributes and physiological reactions, culminating in more tailored and efficacious approaches to health enhancement and wellness management. This initiative acts as a cornerstone in unlocking the full potential of data-driven insights for optimizing individual welfare and societal health.

# Calories Burn Prediction

ORIGINALITY REPORT

| 13% | 8% | 1% | 8% |
|---|---|---|---|
| SIMILARITY INDEX | INTERNET SOURCES | PUBLICATIONS | STUDENT PAPERS |

PRIMARY SOURCES

| | | |
|---|---|---|
| 1 | www.coursehero.com<br>Internet Source | 2% |
| 2 | Submitted to The Robert Gordon University<br>Student Paper | 1% |
| 3 | Submitted to Botswana Open University<br>Student Paper | 1% |
| 4 | Submitted to Brunel University<br>Student Paper | 1% |
| 5 | www.questionpro.com<br>Internet Source | 1% |
| 6 | proceedings.stis.ac.id<br>Internet Source | 1% |
| 7 | Submitted to Robert Kennedy College AG<br>Student Paper | 1% |
| 8 | Submitted to University of North Texas<br>Student Paper | 1% |
| 9 | Submitted to University of West London<br>Student Paper | 1% |

| | | |
|---|---|---|
| 10 | www.relataly.com<br>Internet Source | 1 % |
| 11 | Submitted to Indian School of Business<br>Student Paper | 1 % |
| 12 | Rajeev Joshi, Raaga Sai Somesula, Srinivas Katkoori. "Edge-Driven Intelligence: A Hybrid Machine Learning Strategy for IoT Edge Nodes", 2023 First International Conference on Cyber Physical Systems, Power Electronics and Electric Vehicles (ICPEEV), 2023<br>Publication | 1 % |
| 13 | www.frontiersin.org<br>Internet Source | 1 % |
| 14 | careers.rhimagnesita.com<br>Internet Source | <1 % |
| 15 | media.proquest.com<br>Internet Source | <1 % |

| | | | |
|---|---|---|---|
| Exclude quotes | Off | Exclude matches | Off |
| Exclude bibliography | Off | | |