DATA ANALYTICS
BOOTCAMP
at atomcamp

STEPS AHEAD:

UNLOCKING THE PATTERNS IN AEROFIT TREADMILL SALES

Prepared by :
Hassaan Malik

Prepared for :
Portfolio Project: EDA

Steps Ahead: Unlocking the Patterns in Aerofit Treadmill Sales[1]

Executive Summary

To identify potential patterns in sales of Aerofit's three treadmills on offer, the marketing team conducted a consumer analysis of 180 purchases during the last three months of the outgoing quarter. The team found that 88% of customers are in their 20s and 30s. Moreover, there is a positive relationship between income level and treadmill choice, with middle-income customers purchasing KP281 and KP481, whereas KP781 is the sole choice of high-income customers. The choice of treadmill is also linked to customers' psychographics: the higher the customers rate their fitness level (scale of 1-5), the more likely they are to purchase KP781 (72.5% of customers). There is a significant association among gender and treadmill types. Men are more likely to rate their fitness 5, and consequently, purchase KP781. Considering these key insights, this report makes recommendations for new metrics for data collection, dropping metrics which did not yield actionable insights, tailoring copywriting and advertisement towards specific segments, and collaboration with fitness groups and individuals.

[1] The analysis was conducted using Python on Colab Notebook. To view the code, please follow https://colab.research.google.com/drive/1dCXkUbJWwF8vgGWpEEoWs2GDHkNOniIh?usp=sharing

Introduction

Aerofit is a startup that produces workout equipment, with treadmills being its primary product. It currently offers three treadmills with varying features and costs: The KP281 is an entry-level treadmill that sells for $1,500; the KP481 is for mid-level runners and sells for $1,750; the KP781 treadmill has advanced features, and it sells for $2,500.

Aerofit sells both online and in-store. Its highly trained staff specializes in assisting customers by understanding their needs, making informed suggestions, and allowing them product testing before purchasing (only available in-store).

Data-Driven Insights

Like any modern startup, Aerofit believes in the power of data. Not only does it focus on demographic data of its customer base, but it also seeks to understand their psychographics. By understanding their planned usage, self-perception of their fitness, and goals towards which they are working, it offers professional guidance to suit individual needs. Moreover, customers' insights contribute to the R&D.

Dataset Overview

	Features	Total Observations	Unique Observations	Missing Values	Duplicates/ Inconsistent Data
	Product	180	3	0	0
Demographics	Age	180	32	0	0
	Gender	180	2	0	0
	Education	180	8	0	0
	Marital Status	180	2	0	0
	Income	180	62	0	0
	Income Level*	180	3	0	0
	Age Labelled*	180	3	0	0
	Miles Labelled*	180	4	0	0
Psychographics	Fitness	180	5	0	0
	Usage	180	6	0	0
	Miles	180	37	0	0

** Added during analysis*

Outliers Treatment

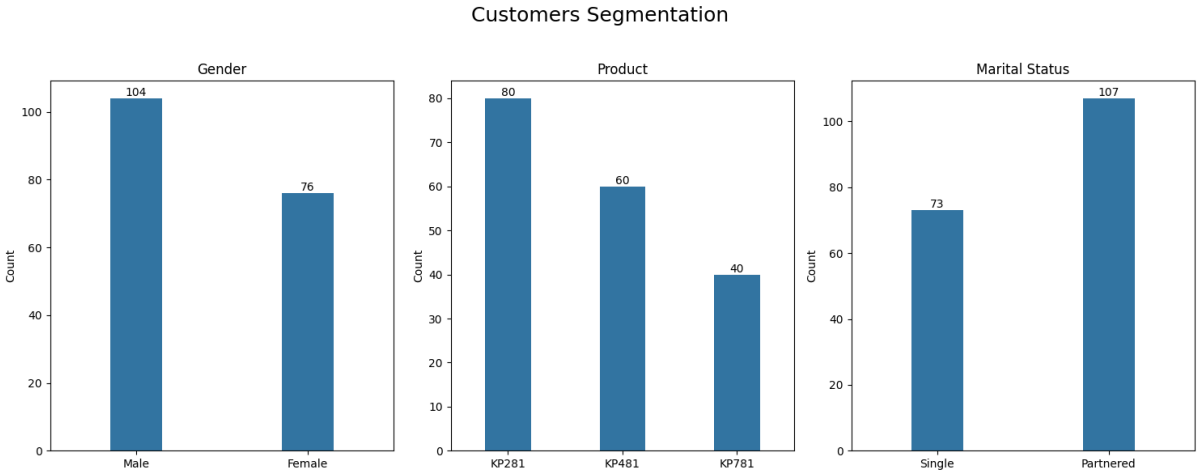
Tukey's interquartile range (IQR) method for outlier detection returned 13 outliers. Features Age, Education, Usage, Fitness, Income, and Miles carried outliers. However, following detailed analysis (see appendix 1), it was observed that only ***Miles*** feature had outliers of concern, which was self-reported by people who deemed their fitness excellent (level 5). Hence, we did not remove the outliers.

Analysis and Key Insights

Highest Selling Treadmill

KP281 is the highest selling product among the three treadmills currently on offer,

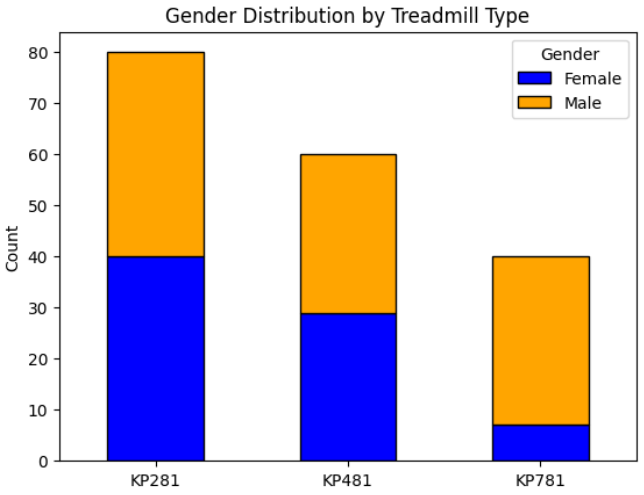
accounting for over 44% of all sales during the last quarter. Its customers may further be divided into demographics such as gender and income level, as well as psychographics such as self-reported fitness level and intended usage per week.



Gender – Product

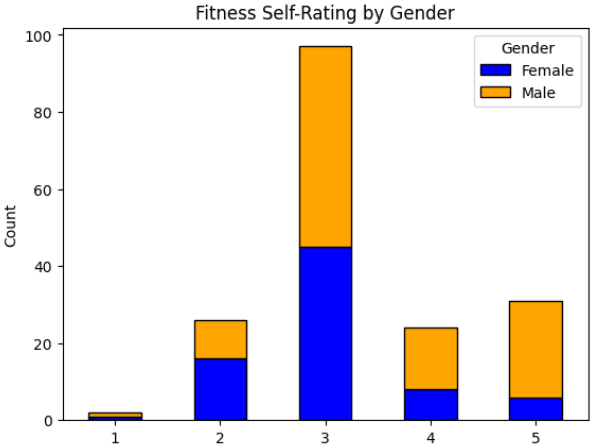
A significant association was found among gender and product type. To investigate the association further, sales of three treadmills were considered against genders.

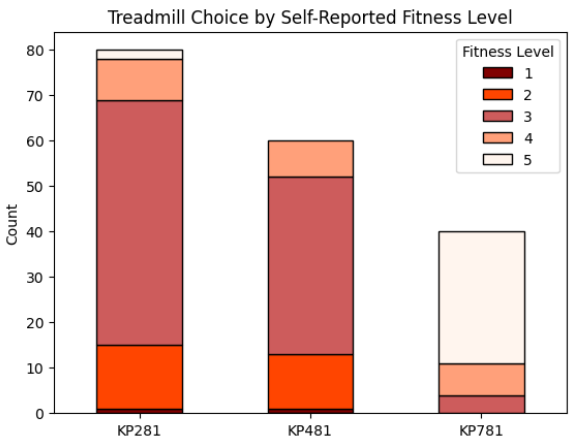
```
Chi-Square Test for Gender vs Product:
-------------------------
Chi-Square Statistic: 12.923836032388664
-------------------------
p-value: 0.0015617972833158714
-------------------------
Significant association found.
```



Gender – Fitness

It is interesting to note that gender and perceived fitness levels affect the choice of treadmill. Female customers were found conservative in self-estimation of their fitness levels, whereas male customers are relatively more liberal in raking their

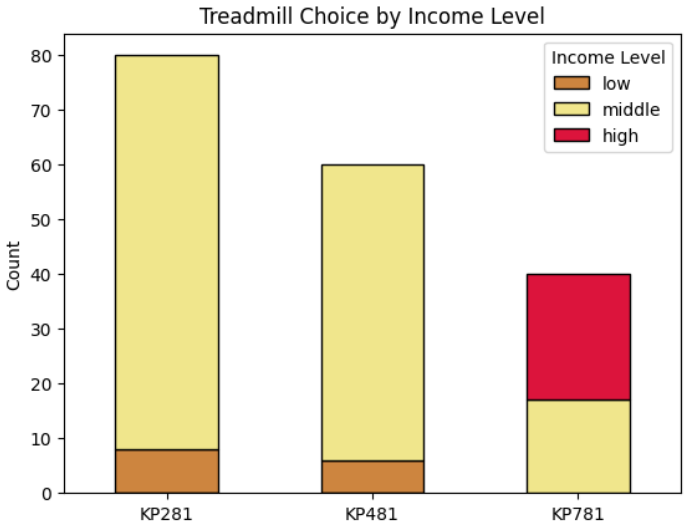




fitness higher on a scale of 1-5. This partially explains why male customers are more likely to buy advanced KP781.

Income – Product

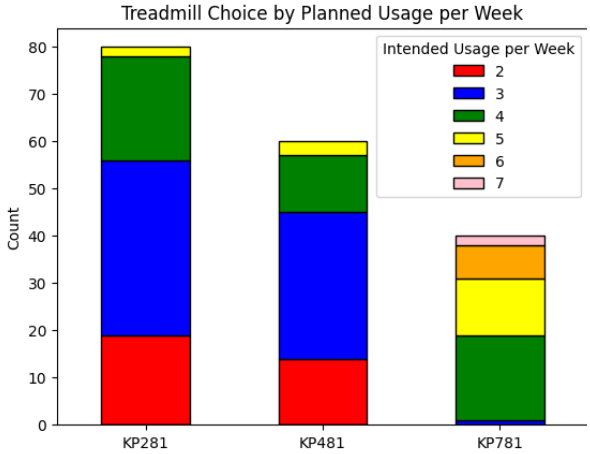
Income level, collected as numeric data but labelled for convenience, showed a positive relationship with prices of treadmills on offer. High-income individuals are more likely to buy KP781 with advanced features, whereas low-income customers are more likely to stick to the entry-level KP281. The interesting thing to note here is the distribution of middle-income customers here. Current marketing appeals to customers' psychographics, i.e., miles to walk/run, which may need to be adjusted for middle-income customers to entice them to consider KP481 and KP781 over KP281.

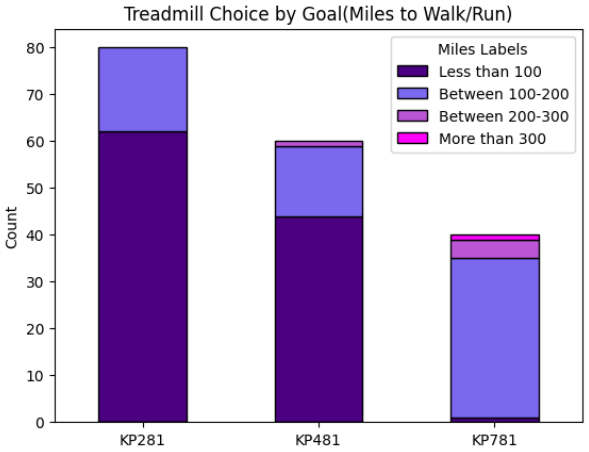


IncomeLevel	low	middle	high
Product			
KP281	8	72	0
KP481	6	54	0
KP781	0	17	23

Product – Usage and Goals

Customers' self-estimation of where they are presently and where they wish to be guides their choices of treadmills. In addition to a positive relationship between perceived fitness level and treadmill of choice, motivated runners are more likely to spend more if they perceive value-to-money in higher-end products like KP781. Comparatively, it is likely that beginner or mid-level runners are comfortable with average features offered by KP281 and KP481. The missing data point that should be collected in the next quarter should be fitness goals. Understanding why

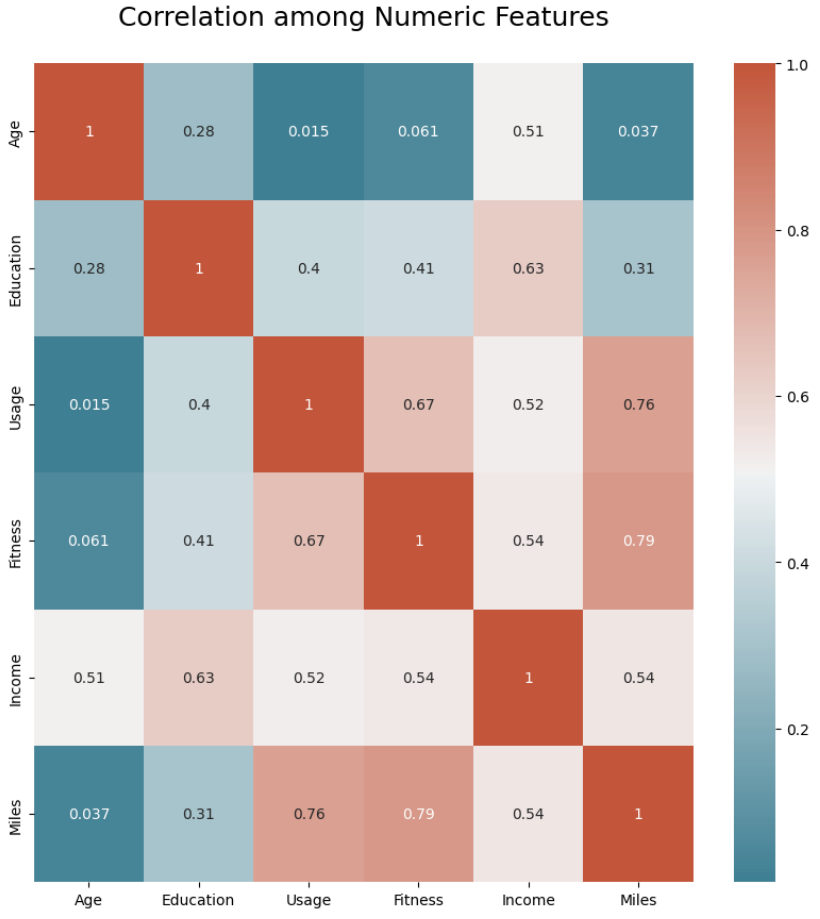




individuals may want to run less than 100 miles or more than 300 miles per week can help Aerofit tailor its marketing campaigns.

Other Relationships of Interest

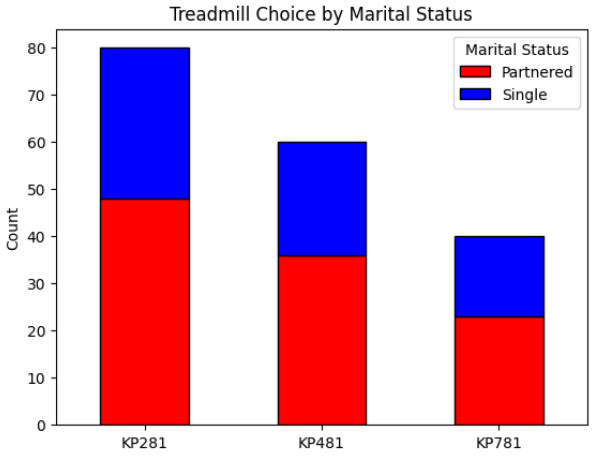
Correlation	Features	Insight	Suggestions
High Positive Correlation	Fitness and Miles (0.79)	This suggests that individuals who self-reported themselves physically fit are more confident of their ability to walk/run long distances.	Data on personal fitness goals will help see if individuals with long distance exercise are training for specific events, which may help identify potential audience and direct advertisement.
	Miles and Usage (0.76)	This relationship is in line with the one above i.e., higher the planned miles, higher the usage and training.	Aerofit can offer tailored workout plans by collaborating with individual physical and marathon trainers.
Moderate Positive Correlation	Fitness and Usage (0.67)	This suggests that individuals who are more physically fit are surer of their frequent use of the treadmill.	Collaborating with fitness influencers may be considered for lead generation.
	Education and Income (0.63)	This suggests that highly educated individuals have jobs in high-paying sectors.	Corporations are a potential customer in that they may be offered tailored packages to add gym rooms in their offices.
Weak Positive Correlation	Age and Fitness (0.061)	This indicates that aged individuals rate themselves lower on fitness levels.	Marketing funds should consider the store visits, telephonic inquiries, online lead generation and conversion from aged individuals to rethink allocation of advertising funds.
	Age and Usage (0.015)	This suggests that aged individuals are less likely to use the treadmill.	



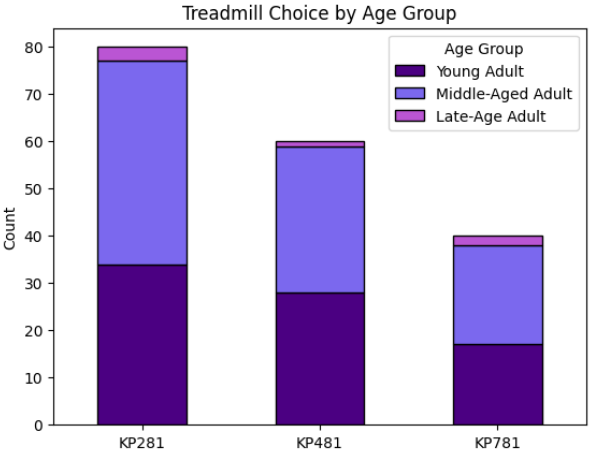
Dropping Data Collection for Insignificant relationships

Marital status did not have any meaningful association with choice of treadmill. So was the case with age groups and products.

```
Chi-Square Test for AgeLabeled vs Product:
-------------------------
Chi-Square Statistic: 1.0621752165223193
-------------------------
p-value: 0.9002261492105591
-------------------------
No significant association found.
```




```
Chi-Square Test for MaritalStatus vs Product:
-------------------------
Chi-Square Statistic: 0.0806554858532839
-------------------------
p-value: 0.9604745988058153
-------------------------
No significant association found.
```



Recommendations

1. Improved Data Collection

a) More features should be added to collect data regarding more relevant metrics of demographics–e.g., purchase method (online vs. In-store), profession, social media use etc.–and psychographics–fitness goals (general endurance, weight loss etc.), tech preference.

b) Aerofit should develop online communities to stay connected with its customers to collect feedback and suggestions regarding new features and improvements.

c) While age may become useful with improved demographic and psychographic data coming in, marital status data is not expected to yield any meaningful insights and should be replaced with better metrics.

2. Middle-Income Group

The middle-income group has the most diversified consumer pattern. More nuanced data can help understand this segment better.

3. Tailor Copywriting and Advertisement

Current advertisements portray all treadmills based on presumed

psychographics of customers (self-perceived beginners, novice, and professionals). It is recommended that:

1. For KP281, copywriting and marketing should focus on demographics (gender and income level).
2. For KP781, likened by the high-income individuals and among them, men, copywriting should give psychographics more importance.

4. Collaboration with Fitness Communities and Influences

For customers of KP481 and KP781, after understanding their specific goals, Aerofit can offer tailored workout and meal plans by collaborating with fitness groups and individual physical trainers.

5. Value-for-Money Advertisements

Aged individuals (+45) are less likely to buy treadmills, therefore, physical and online advertisement should look for an alternative target audience. A good consideration would be startups and corporate offices that may be offered to add gyms for employees.

Conclusion

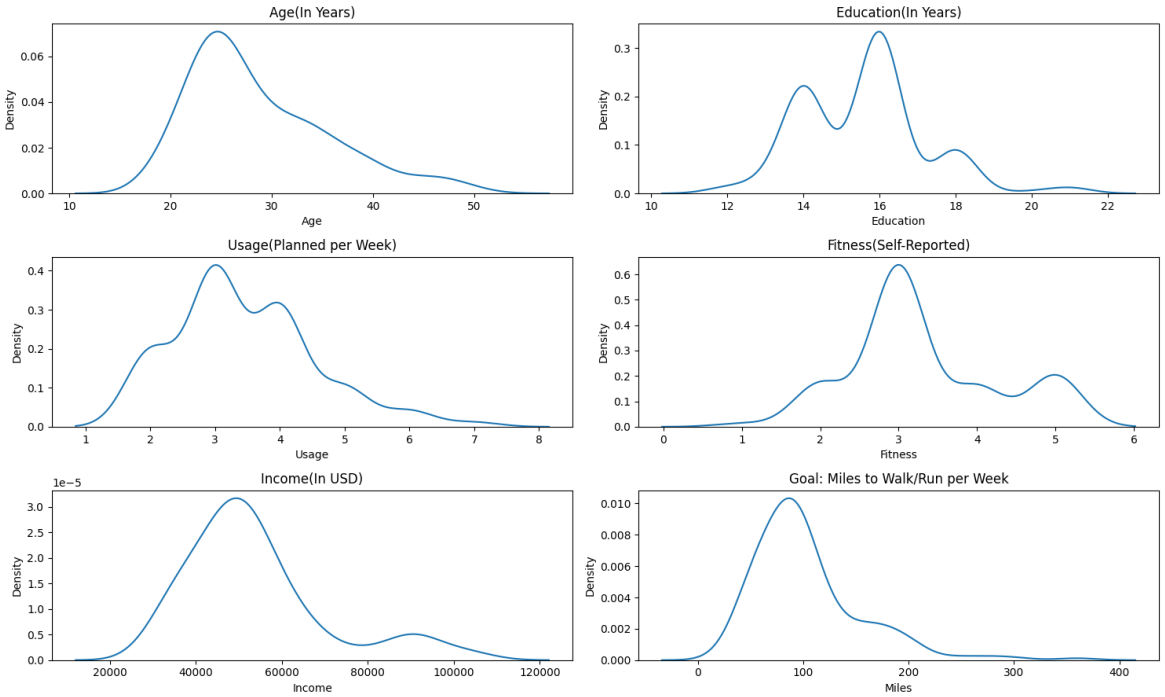
This report has demonstrated that important differences exist in Aerofit's customers' purchasing patterns for its three treadmills KP281, KP481, and KP781 on offer. The study undertook a detailed consumer analysis to show that both demographics (gender and income-level) and psychographics (perceived fitness, usage plans, and running/walking goals) influence the latter's choices. A nuanced understanding not only highlighted potential metrices for which data needs to be collected for better insights (fitness goals, tech preference, social media use etc.), it also showed that certain metrices may yield better insights (age group, marital status) vis-à-vis new data metrices.

Despite the study's limitations, including metrices which did not provide meaningful insights, these findings provide a strong foundation for future data collection, tailoring of marketing strategies, lead generation, conversions, and sales.

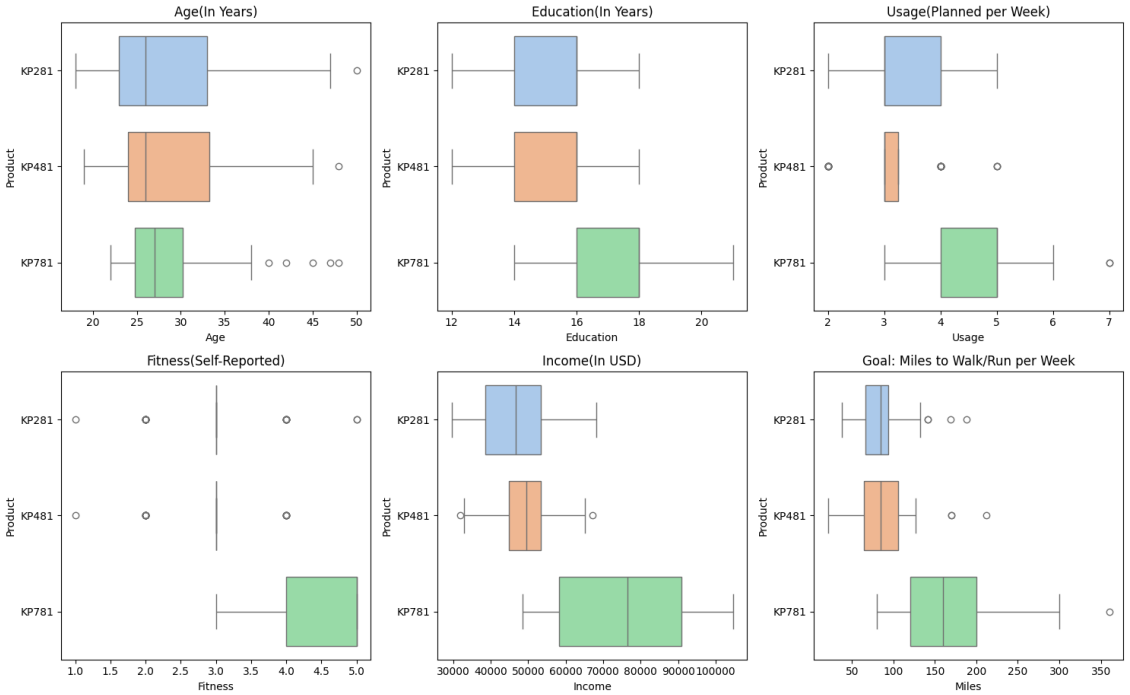
In conclusion, Aerofit's current sales funnel for its three treadmills has performed quite well, establishing a strong foundation for scaling and improvements.

Appendix 1: Dealing with outliers

The provided data set was free of missing or duplicated data, however, none of the numeric features were anywhere near normally distributed. The Tukey IQR test was carried out to detect 13 outliers. However, after the dropping operation was undertaken, 9 out of 13 outliers persisted. As distribution was not normal, neither standard deviation, nor z-score or modified z-score method could be carried out.



Distribution and Outliers in Demographics and Psychographics



Hence, a manual investigation was undertaken by comparing the numeric features before and after removal of outliers.

```
# before droppping outliers
af_treadmill.describe()
```

	Age	Education	Usage	Fitness	Income	Miles
count	180.000000	180.000000	180.000000	180.000000	180.000000	180.000000
mean	28.788889	15.572222	3.455556	3.311111	53719.577778	103.194444
std	6.943498	1.617055	1.084797	0.958869	16506.684226	51.863605
min	18.000000	12.000000	2.000000	1.000000	29562.000000	21.000000
25%	24.000000	14.000000	3.000000	3.000000	44058.750000	66.000000
50%	26.000000	16.000000	3.000000	3.000000	50596.500000	94.000000
75%	33.000000	16.000000	4.000000	4.000000	58668.000000	114.750000
max	50.000000	21.000000	7.000000	5.000000	104581.000000	360.000000

```
# after dropping outliers
af_outliers_removed.describe()
```

	Age	Education	Usage	Fitness	Income	Miles
count	167.000000	167.000000	167.000000	167.000000	167.000000	167.000000
mean	28.377246	15.407186	3.317365	3.191617	50865.802395	94.640719
std	6.730274	1.473279	0.957468	0.884499	13306.883247	38.622395
min	18.000000	12.000000	2.000000	1.000000	29562.000000	21.000000
25%	23.500000	14.000000	3.000000	3.000000	43206.000000	66.000000
50%	26.000000	16.000000	3.000000	3.000000	50028.000000	85.000000
75%	33.000000	16.000000	4.000000	3.500000	54678.500000	109.000000
max	50.000000	21.000000	7.000000	5.000000	104581.000000	212.000000

It was observed that only Miles had outliers of interest. After locating–loc()–for indices of outliers, a pattern emerged: usage, fitness, and miles, all self-reported had extreme values, hence none of them can be classified as data entry error. It was for this reason that outliers were not removed.

	Product	Age	Gender	Education	MaritalStatus	Usage	Fitness	Income	Miles
155	KP781	25	Male	18	Partnered	6	5	75946	240
166	KP781	29	Male	14	Partnered	7	5	85906	300
167	KP781	30	Female	16	Partnered	6	5	90886	280
170	KP781	31	Male	16	Partnered	6	5	89641	260
173	KP781	35	Male	16	Partnered	4	5	92131	360