

Spending Habits

Group Nova

Emily Nguyen, Voldie Louza, Ben Elhadad, Kaylynn Francisco-Nelson

Introduction

We investigated the leisure spending habits of individuals based on a multitude of variables including age, gender, and the type of spending they do the most. We brainstormed various independent variables that could have an effect on leisure spending like marital status, family obligation, and transportation. However, we landed on age, gender, spending type, and monthly income. We thought it would be interesting to see how spending habits differed between different types of people. This investigation is important for understanding demographic spending trends.

Our hypothesis is that **age, gender, employment status, and monthly income do not significantly influence leisure spending**. This research can reveal how these variables shape spending patterns and whether they align with or deviate from traditional economic theories.

Data

We collected our data through random surveying. We aimed to get an unbiased sample by sending our survey to people of various ages. This goal was met by sending it to our peers, family members, and friends. The survey included questions on demographics (age, gender, and employment), monthly income, and leisure spending habits, such as expenditure categories.

When it comes to cleaning our data, we found that with our survey many people were inputting character values with their open responses. For example, when we asked for income or the average amount spent per month, some responses added commas or dollar signs. So we had to go through and make sure all of the observations only had numbers and no characters that might cause errors in our analysis.

To make our data usable, we went through and made the variables that were categorical into binary representations. This includes:

- **Age** was categorized into groups: 18–24, 25–31, 32–38, 39–45, and 45+.
- **Gender** was encoded as binary: 1 for females, and 0 for males.
- **Employment Status**: 1 for employed, 0 for unemployed.
- **Spending Types**: Categorized into binary variables for spending on eating out, hobbies, leisure activities, shopping, subscriptions, technology, and travel.

Exploratory Analysis

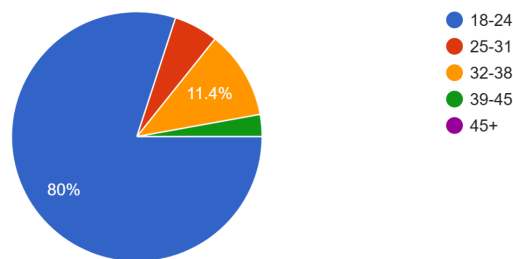
Dataset Overview

The cleaned dataset includes the following variables:

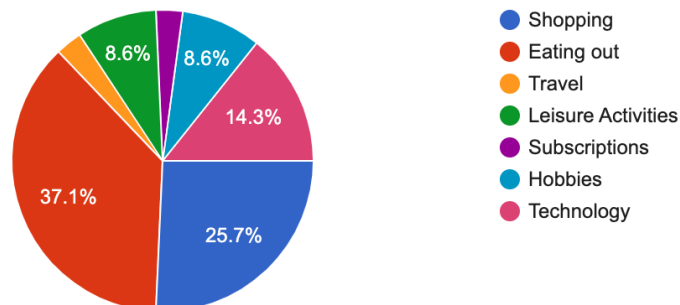
- **Independent Variables:** Age groups, gender, employment status, and monthly income.
- **Response Variable:** Leisure spending amount.

Key Observations

1. **Age Distribution:** The majority (80%) of respondents were aged 18–24, indicating potential sample bias.

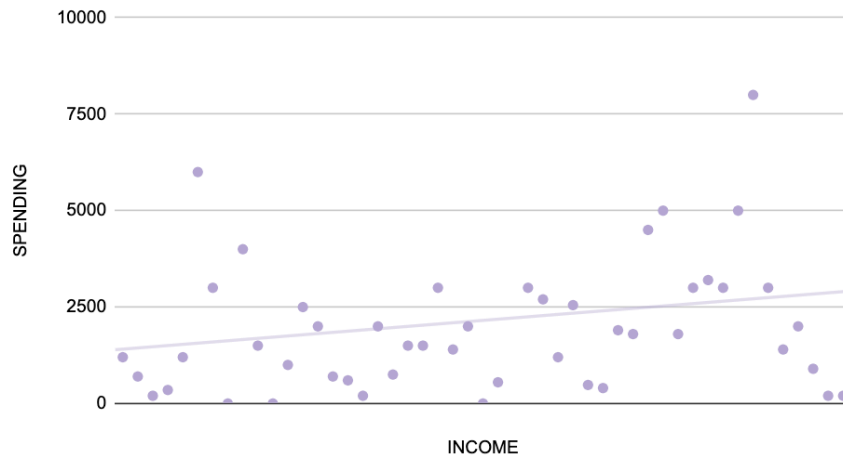


2. **Spending Trends:** We found that most of our respondents spent their leisure money on either eating out or shopping.



3. **Income Variability:** After creating a scatter plot to analyze the relationship between spending and income, we found that monthly income ranged widely, but higher income did not correspond to higher leisure spending consistently. We see that there is an overall trend of leisure spending increasing with an increase in income. However, on the most-right side of the scatter plot, we discovered that many respondents did not spend a significant amount of money relative to those who have a lower income.

SPENDING vs. INCOME



Sample Visualization

Below is an example table showing the first 10 rows of the dataset:

The SAS System																	
Obs	Age18_24	Age25_31	Age32_38	Age39_45	Age45PLUS	Gender	Employment	Eating_Out	Hobbies	Leisure_Activities	Shopping	Subscriptions	Technology	Travel	Income	Budget	Spending
1	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	500
2	1	0	0	0	0	1	1	0	0	0	1	0	0	0	1200	1	300
3	1	0	0	0	0	0	0	1	0	0	0	0	0	1	700	0	150
4	1	0	0	0	0	0	0	1	1	0	0	0	0	0	200	1	50
5	1	0	0	0	0	0	0	0	0	0	1	0	0	0	350	1	200
6	1	0	0	0	0	0	0	1	0	0	0	0	1	0	1200	1	250
7	0	0	1	0	0	0	1	0	0	0	0	0	1	0	6000	0	500
8	1	0	0	0	0	0	0	1	1	0	0	0	0	0	3000	0	200
9	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	400
10	0	0	1	0	0	0	0	1	0	1	0	0	0	0	4000	0	500

Analysis

To identify predictors of leisure spending, we performed a multiple linear regression. The initial model included all of the variables from our survey, but we adjusted it later on to age groups, gender, employment status, and monthly income.

Variable Refinement:

After performing a regression analysis, we found that:

- The variable **Age 45+** was found to be statistically insignificant ($p\text{-value} > 0.05$) and was dropped from the model. This was understandable as our dataset didn't have any responses from people aged 45 or over.

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	B	1094.72445	379.41288	2.89	0.0067
Age18_24	B	-484.77695	309.21134	-1.57	0.1262
Age25_31	B	-177.34906	322.51001	-0.55	0.5860
Age32_38	B	1.79603	255.50230	0.01	0.9944
Age39_45	B	-40.71678	438.77340	-0.09	0.9266
Age45PLUS	0	0	.	.	.
Gender	1	30.50026	97.59888	0.31	0.7566
Employment	1	17.84385	104.81557	0.17	0.8658
Eating_Out	B	-346.37101	235.10336	-1.47	0.1499
Hobbies	B	-716.25386	285.81368	-2.51	0.0172
Leisure_Activities	B	-123.86447	303.34787	-0.41	0.6856
Shopping	B	-287.01414	235.70405	-1.22	0.2317
Subscriptions	B	-449.78062	336.32207	-1.34	0.1900
Technology	B	-267.62488	265.56676	-1.01	0.3207
Travel	0	0	.	.	.
Income	1	-0.00742	0.04234	-0.18	0.8619
Budget	1	-25.90887	85.29047	-0.30	0.7632

- Variables representing specific spending categories (e.g., hobbies, subscriptions, travel) were also evaluated. These variables showed weak significance and were removed from the model due to their limited alignment with our core research question. So we decided to exclude these variables to improve the clarity and precision of our model while ensuring that the retained variables provide more meaningful contributions. This step helped to reduce potential multicollinearity and improve our model.

Key Findings

- After performing a stepwise regression, only **Age 18–24** and **hobbies** spending were identified as significant predictors. However, this result was heavily influenced by sample bias, as 80% of respondents fell within the 18–24 age group.

Variable Hobbies Entered: R-Square = 0.3519 and C(p) = -5.0996

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	1421968	710984	12.49	<.0001
Error	46	2618566	56925		
Corrected Total	48	4040535			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	694.36861	66.65805	6177045	108.51	<.0001
Age18_24	-375.09195	77.47218	1334411	23.44	<.0001
Hobbies	-336.85329	112.49555	510407	8.97	0.0044

We decided to move on by testing all of the age ranges, except age 45+ because without all of the categorical age predictors, our results would not capture spending habits of multiple age ranges.

- Final Model and Results: The refined regression model showed that all remaining predictors (age groups, gender, employment status, and income) had **p-values > 0.05**. This could indicate that there is no significant relationship with leisure spending. Additionally, our model's p-value is 0.0514 which is greater than our significance level at 0.05. Therefore, we fail to reject the null hypothesis that these variables do not significantly affect leisure spending.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	1113532	159076	2.23	0.0514
Error	41	2927003	71390		
Corrected Total	48	4040535			

Root MSE	267.18966	R-Square	0.2756
Dependent Mean	408.16327	Adj R-Sq	0.1519
Coeff Var	65.46147		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	337.45813	217.05263	1.55	0.1277
Age18_24	1	8.04983	217.61291	0.04	0.9707
Age25_31	1	386.57509	240.42970	1.61	0.1155
Age32_38	1	263.78050	230.84361	1.14	0.2598
Age39_45	1	248.49581	410.02217	0.61	0.5478
Gender	1	-9.10906	81.97512	-0.11	0.9121
Employment	1	-54.41530	96.55312	-0.56	0.5761
Income	1	0.00712	0.03984	0.18	0.8590

Interpretation of Results

1. Biases: The overrepresentation of respondents aged 18–24 likely skewed the results, limiting the generalizability of our findings. A more balanced dataset across different age groups would have been more useful and significant to our model.
2. Hypothesis Testing: The lack of significance in almost all of our predictors also indicates that our dataset might lack the variability it needs to find meaningful differences in spending habits.

Conclusion

This study attempted to identify key drivers of leisure spending through a multilinear regression model. However, we found no significant predictors due to sample bias and limited variable diversity. We found that with all the P-values of the remaining variables, we failed to reject the null hypothesis that age, gender, employment, and monthly income do significantly affect leisure spending. This result is most likely biased due to the fact that 80% of the responses were of people between the ages of 18 and 24, which was expected given our data pool.

In the future, we would like to:

1. Expand the sample to include a more balanced representation of age groups and demographics.
2. Investigate other potential predictors, such as geographic location or cultural factors.
3. Conduct stratified sampling to reduce bias and improve generalizability.