**Customer Spending Regression Analysis Report**

**1. Introduction**

**Objective:**  
The purpose of this analysis is to understand customer personality, preferences, and behavior so that the business can:

* Segment its customer base
* Optimize and measure marketing campaigns
* Improve customer retention

To achieve these goals, we examined the drivers of customer spending using both correlation analysis and multiple linear regression.

**2. Data and Variables**

**Dependent Variable:**

* **TotalSpend:** The total amount spent by each customer.

**Independent Variables:**  
We considered a wide range of predictors, including:

* **Demographic and Socioeconomic:** Age, Income, TotalChildren
* **Behavioral:** Recency (days since last purchase), TotalAcceptedCmp (number of campaigns accepted), NumCatalogPurchases, NumWebPurchases, NumStorePurchases, NumWebVisitsMonth
* **Categorical Variables:** Education (converted to dummy variables: Graduation [reference], Master, Basic, 2nCycle, PhD) and Marital\_Status (converted to dummy variables with one reference category)
* **Service Quality:** Complain (whether the customer made a complaint)

**3. Correlation Analysis**

Before running the regression, we calculated the Pearson correlation coefficients to understand the relationships between each predictor and TotalSpend.

**Key findings include:**

* **Strong Positive Relationships:**
  + Income (0.819): Higher-income customers tend to spend more.
  + NumCatalogPurchases (0.779) and NumStorePurchases (0.675): Frequent purchases via catalog or stores strongly associate with higher spending.
  + TotalAcceptedCmp (0.456): Customers accepting more promotional campaigns tend to spend more.
* **Moderate to Strong Negative Relationships:**
  + TotalChildren (-0.499): More children are linked to lower spending, possibly due to additional financial commitments.
  + NumWebVisitsMonth (-0.500): Higher website visits (without a matching increase in purchases) associate with lower TotalSpend.
* **Weaker Relationships:**
  + Age (0.111): Weak positive correlation.
  + Several education and marital status dummies show weak or non-significant associations, suggesting they may not be strong independent predictors.

| **Variable** | **Correlation with TotalSpend** | **Note (Statistical Significance)** |
| --- | --- | --- |
| **Age** | 0.111\*\* | Weak positive correlation |
| **Education\_Graduation** | 0.024 | Weak, likely not significant |
| **Education\_Master** | 0.004 | Weak, not significant |
| **Education\_Basic** | -0.137\*\* | Weak negative correlation |
| **Education\_2nCycle** | -0.057\*\* | Weak negative correlation |
| **Education\_PhD** | 0.058\*\* | Weak positive correlation |
| **Marital\_Status\_Single** | -0.001 | Not significant |
| **Marital\_Status\_Married** | -0.020 | Not significant |
| **Marital\_Status\_Widow** | 0.042\* | Weak positive correlation |
| **Marital\_Status\_Unknown** | 0.004 | Not significant |
| **Marital\_Status\_Divorced** | 0.003 | Not significant |
| **Income** | 0.819\*\* | Very strong positive correlation |
| **Recency** | 0.020 | Not significant |
| **TotalChildren** | -0.499\*\* | Moderate negative correlation |
| **NumDealsPurchases** | -0.065\*\* | Weak negative correlation |
| **NumWebPurchases** | 0.520\*\* | Moderate positive correlation |
| **NumCatalogPurchases** | 0.779\*\* | Strong positive correlation |
| **NumStorePurchases** | 0.675\*\* | Strong positive correlation |
| **NumWebVisitsMonth** | -0.500\*\* | Moderate negative correlation |
| **Complain** | -0.093\*\* | Weak negative correlation |
| **TotalAcceptedCmp** | 0.318\*\* | Moderate positive correlation |

**Interpretation:**  
The correlation analysis helped identify variables with strong associations to spending, informing variable selection for the regression model. Key predictors like Income, purchase behaviors, and family size were prioritized, while variables with weaker correlations were flagged for potential exclusion.

**4. Multiple Linear Regression Analysis**

**4.1 Model Specification**

Based on the correlation results and business rationale, we constructed a multiple linear regression model with TotalSpend as the dependent variable and a selection of independent variables:

* **Included variables:** Age, Income, Recency, TotalChildren, TotalAcceptedCmp, NumCatalogPurchases, NumWebPurchases, NumStorePurchases, NumWebVisitsMonth, and dummy variables for Education (reference: Graduation).
* **Excluded variables:** Non-significant variables (e.g., some education levels, marital status dummies, and Complain) were removed to simplify the model.

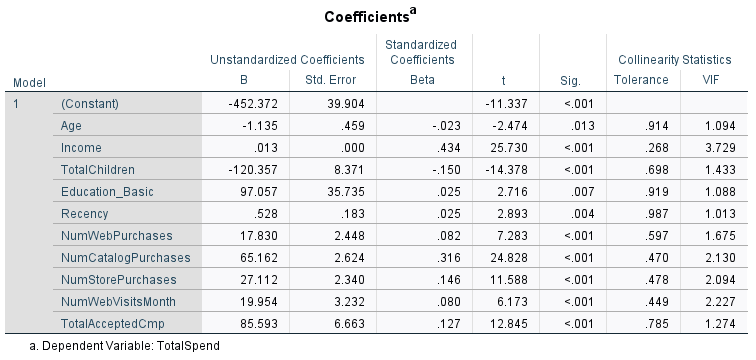
**4.2 Model Results**

The final regression model yielded an R² = 0.829, indicating that 82.9% of the variance in TotalSpend is explained by the predictors, reflecting a strong model fit.

| **Predictor** | **Coefficient (B)** | **Significance (p-value)** | **Interpretation** |
| --- | --- | --- | --- |
| Income | +0.013 | < 0.001 | Strong positive effect: higher income → more spending. |
| TotalAcceptedCmp | +85.593 | < 0.001 | More campaigns accepted → significantly more spending. |
| NumCatalogPurchases | +65.162 | < 0.001 | More catalog purchases → higher spending. |
| NumStorePurchases | +27.112 | < 0.001 | More in-store purchases → increased spending. |
| NumWebPurchases | +17.830 | < 0.001 | More web purchases → increased spending. |
| TotalChildren | -120.357 | < 0.001 | More children → lower spending. |
| Recency | +0.528 | 0.004 | Recent purchases slightly increase spending. |
| Age | -1.135 | 0.013 | Older customers tend to spend slightly less. |
| NumWebVisitsMonth | +19.954 | < 0.001 | More website visits correlate with increased spending. |
| Education\_Basic | +97.057 | 0.007 | Compared to Graduation, Basic education → higher spending. |

**Note on Reference Categories:**

* Education: Education\_Graduation is the reference. Coefficients indicate differences relative to this group.
* Marital\_Status: One category (e.g., Married) is the reference; its coefficient is not directly shown.



**5. Linking Correlation and Regression Results**

* Variables with strong correlations (Income, Purchase Behaviors, TotalAcceptedCmp) also show significant regression coefficients.
* The negative correlation of TotalChildren aligns with its significant negative effect in the regression.
* Weak or non-significant variables from the correlation analysis were excluded from the final model.
* The high R² confirms that the selected predictors capture most of the variation in customer spending.

This confirms that our model selection based on correlation analysis was effective, yielding a robust predictive model.

**6. Recommendations Based on Analysis**

**1. Target High-Income Customers with Premium Offers:** Leverage their higher spending capacity by designing exclusive campaigns.

**2. Increase Campaign Engagement:** Use personalized messaging and timing to improve campaign acceptance rates.

**3. Strengthen Purchase Channels:** Invest in improving catalog, in-store, and web shopping experiences, focusing on seamless transactions and customer support.

**4. Reactivate Recent Buyers:** Implement loyalty incentives and reminders for customers who purchased recently.

**5. Offer Family-Focused Promotions:** Create budget-friendly bundles and discounts for customers with larger families.

**6. Optimize Digital Interaction:** Enhance the website and online touchpoints to convert more visits into purchases.

**7. Conclusion**

Starting from correlation analysis to a refined multiple linear regression model, this analysis identifies the main factors driving customer spending. By removing weak predictors and focusing on significant ones, we provide actionable insights to support segmentation, campaign optimization, and retention strategies that align with project goals.