JSC370 Final Report

Ella Hugie

I. Introduction

With the continual evolution of digital technology, the advertising landscape across Canada has seen substantial change. Traditional advertising mediums, such as newspapers and radio, are being gradually replaced by newer forms of advertising. This study aims to delve into the effects of digitization on advertising expenditure across Canada, carrying significance for both policymakers and consumers alike. To explore these effects, I analyzed an advertising expenditure dataset from the Government of Canada Open Portal, covering the years 2018 to 2023 which have seen tremendous digitization. This dataset captures over 94% of national agency spending across Canada's media platforms and includes variables such as year, quarter, month, province, market, type of media, media subtype, and ad spending. Through statistical analysis, this study addresses the question: What is the impact of digital technology on advertising dynamics in Canada? Furthermore, is this impact uniform across the country, or are there specific regions exhibiting varying degrees of change?

II. Methods

The dataset used for this analysis was acquired from the Government of Canada Open Portal, which makes comprehensive datasets pertaining to various aspects of Canadian governance and industry publicly available. This data includes comprehensive information on Canadian Agency Advertising Gross Spending from 2018 to 2023, although it should be noted that the 2023 data does not encompass the full calendar year which was taken into account when performing analyses. Following acquisition, the dataset underwent a cleaning and wrangling process which involved checking for missing values, renaming variables, factoring categorical variables, and creating new variables that will be used later in the analysis. Two notable variables that were created were Total Spending and Percentage Digital Spending.

Next, exploratory data analysis was conducted to examine the dataset's underlying distributions, correlations, and outliers. Multiple univariate and bivariate plots, including faceted bar plots and scatterplots were fitted to key variables, such as Ad Spending, to gain insight into underlying relationships. Then, the assumptions of linear regression were verified to ensure the reliability of the linear regression models fit to the data. First, the linearity assumption was checked by plotting residuals against fitted values and looking for any distinct patterns. Then, the residuals were examined for the lack of distinct patterns to ensure independence of errors. Homoscedasticity, the constant variance of errors, was assessed by looking for fanning or other discernible patterns indicating a change in spread in the plot of residuals against fitted values. Finally, the normality assumption was evaluated by plotting a QQ-plot of residuals and looking for how well the data fit to a bell-shaped curve, indicated by the diagonal line in this plot.

After verifying the assumptions of linear regression, the modeling process began. Two different linear regression models were fitted to the data to accomplish different objectives. The first model aimed to predict advertising spending based on a multitude of predictor variables to gain insight into how and where advertising money is allocated. To determine the optimal subset of predictor variables to predict ad spending, both stepwise AIC and stepwise BIC variable selection methods were used. Stepwise AIC and Stepwise BIC iteratively adjust predictors based on their statistical significance as determined by the Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC) respectively. The second model aimed to predict a new variable created called Percentage Digital which calculated the percentage of the total advertising money spent in any particular year on digital advertising. This model was created with the specific goal of quantifying whether or not the advertising landscape is shifting towards digital marketing and its variables were selected manually as opposed to using a variable selection technique. Both models were trained using ordinary least squares, with the objective of minimizing the sum of squared residuals.

After selecting the final models, they were evaluated by several statistical metrics to evaluate their predictive performance. Firstly, the Mean Squared Error (MSE) and Mean Absolute Error (MAE) were calculated. While MSE quantifies the average squared difference between predicted and actual values, giving higher weights to large errors, MAE measures the average absolute difference, providing a more direct interpretation of the model's accuracy. Also, both the R-squared value and Adjusted R-squared value were calculated for each model to assess goodness of fit. R-squared represents the proportion of variance in the outcome variable that is explained by the predictors in the model. Thus, it ranges from 0 to 1, with higher values indicating better fit. While similar, the Adjusted R-squared also adjusts for the number of predictors in the model, penalizing the addition of unnecessary variables.

Following model assessment, each model was used to explore the research question. This included making predictions to see how advertising expenditure will likely continue to evolve in the next five to ten years in Canada. Specifically, the models were used to predict the exact year in which the Canadian advertising budget is expected to become completely overtaken by digital advertising.

III. Results

The exploratory data analysis revealed high completeness and consistency in the dataset. Minimal missing data was observed, and inconsistencies were identified and addressed. Furthermore, all variables appeared to be appropriately distributed, confirming the reliability and suitability of the dataset for further use.

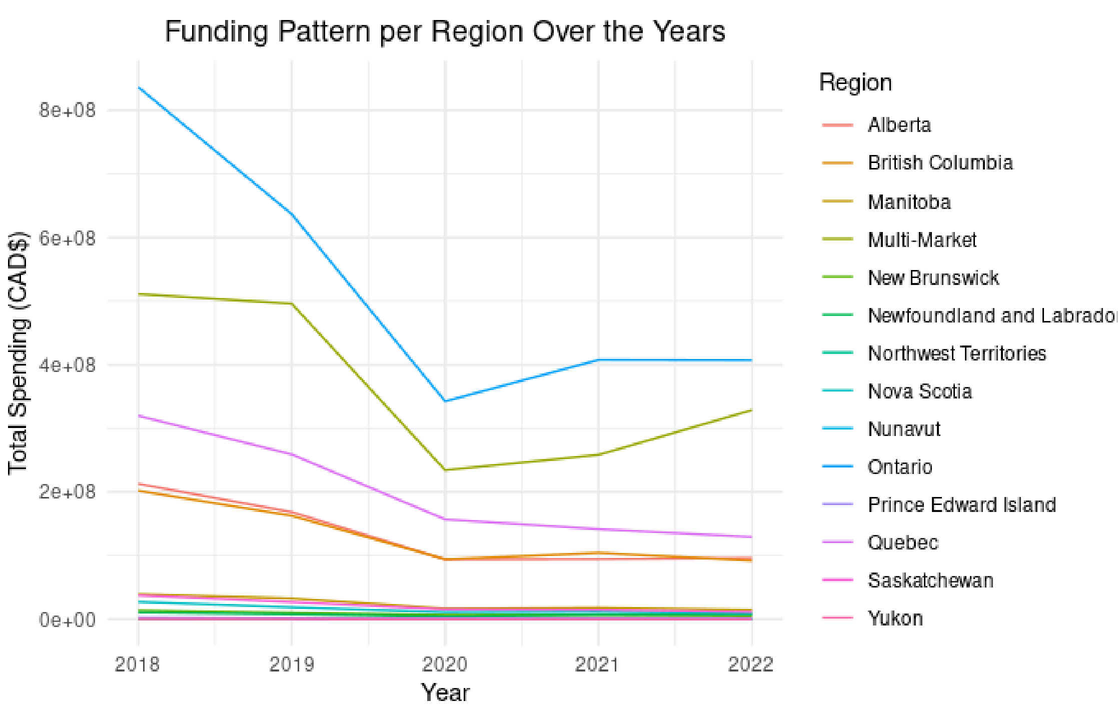

Figure 1: Trends in Funding Allocation per Region Over Time (excluding National)

Preliminary analysis indicated a notable correlation between Ad Spending and Region as displayed in Figure 1 which highlights how region specific advertising funding is decreasing over time. Not shown in this graph is National spending over time as it dwarfs these region-specific trends, however it is steadily increasing over time. This multiple line chart demonstrates that advertising is becoming increasingly holistic in nature and Canada is moving away from region specific methods of advertisements potentially due to the ease and efficacy of national advertising. This chart also displays that advertising was steadily decreasing across the nation until 2020, the year in which the COVID-19 pandemic hit, where it started to plateau or increase in certain provinces. Furthermore, by looking at the total funds spent solely on Radio, Newspaper, and Digital advertising, there are very interesting insights revealing the dynamic of how the advertising landscape is changing. As seen in Figure 2, the percentage of advertising being spent on Digital methods is steadily increasing from 2018 to 2023 while Radio and Newspapers are both steadily decreasing. Note, data was not collected throughout the entire 2023 year, hence the drop in total spending on digital advertising, however the respective digital spending percentage is still increasing.

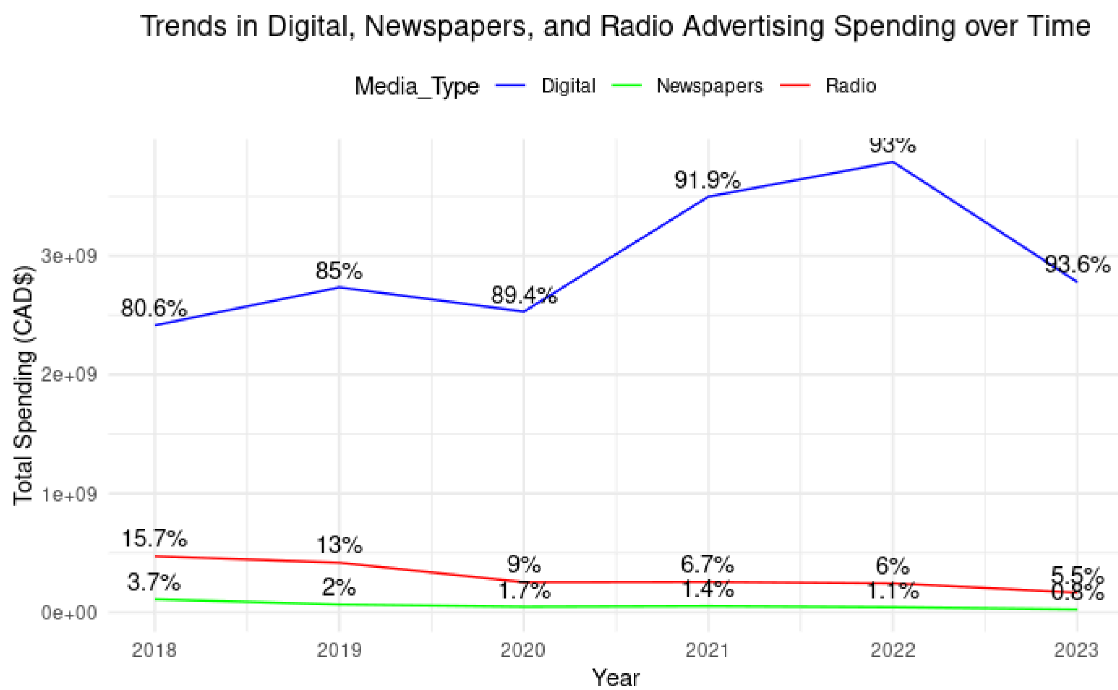

Figure 2: Radio vs. Newspapers vs. Digital Advertising spending Percentages Over Time

Next the models were fit using both the variable selection techniques previously mentioned. The stepwise AIC procedure resulted in a model with "Ad Spending" as the response variable and predictors "Month," "Market," "Media Type," and "Media Sub-type." The BIC procedure yielded a similar model, with the addition of a numeric representation of "Month" labeled as "Month_Numeric." Upon inspection, both models produced practically identical outcomes, as "Month" and "Month_Numeric" represented the same information. Thus, the AIC model was chosen for simplicity. Additionally, the model predicting “Percentage Digital” was fitted manually. Initially, I fit a model incorporating all predictors, but found that none exhibited significant statistical significance except for the variable "Year." Therefore, I opted to simplify the model to include only the predictor "Year." A visual representation of the final two models can be found in Figure 3.

Linear Regression Models

Model Selection Technique	**Response**	**Predictors**	**MSE**	**MAE**	R^2	**Adj.** R^2
Stepwise AIC	Ad Spending	Month, Market, Media Type, Media Sub-type	2.48×10^{13}	1782111	0.66	0.66
Manual	Percentage Digital	Year	1.83	1.25	0.96	0.96

Figure 3: Performance Metrics of both models predicting Total Ad spending and Digital Advertising Percentage

While the Stepwise AIC Model exhibited a high Mean Squared Error (MSE) and Mean Absolute Error (MAE), it's important to note that these values are in units of dollars. Therefore, in the context of the massive scale of the advertising budget, these errors may not be considered unreasonably high. The R-squared value was 0.66, with an Adjusted R-squared value of 0.66. These metrics suggest that while the stepwise AIC model captured some variation in Ad Spending, there were limitations in its predictive accuracy. Despite the manually-selected Percentage Digital model's simplicity, this model achieved a very low MSE of 1.83 and MAE of 1.25. This MAE indicates that, on average, the model's predictions for the percentage of the total advertising budget spent on digital advertising deviates by only approximately 1.25 percentage points from the actual value. Moreover, the R-squared and Adjusted R-squared values were both very high at 0.96, suggesting that the model explains a substantial portion of the variance in Percentage Digital. Note, both models have the same R-squared and Adjusted R-squared values respectively, indicating that the predictors already included in the models are the most influential factors in explaining the variability in the response variable.

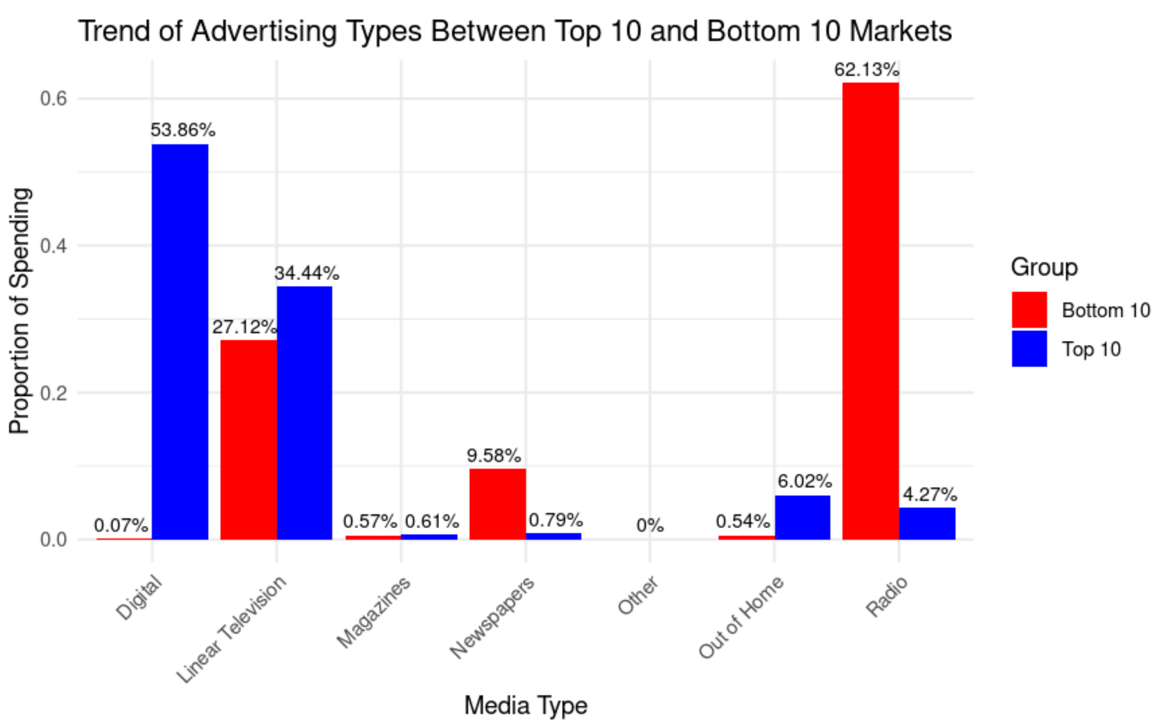

Figure 4: Bar Plot of Proportion of Spending per Media Type, colored by Top 10 and Bottom 10 markets

Finally, the Stepwise AIC model's inclusion of the Market variable prompted further exploration into how advertising type varies across markets. To further explore this, the Proportion of Spending against Media Type was plotted, colored by the top and bottom 10 markets determined by size. As seen in Figure 4, the top 10 markets advertising seems to be majority Digital, Television, and Out of Home. In contrast, the bottom 10 markets advertising seems to be majority Television, Radio, and Newspapers.

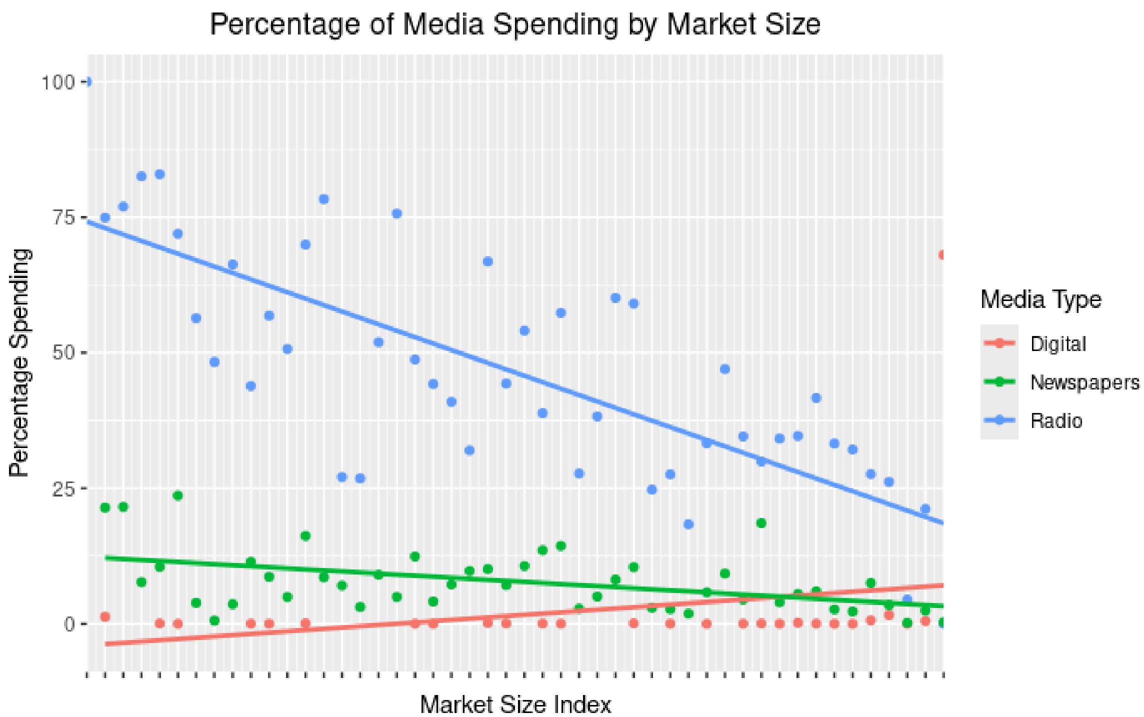

Figure 5: Multiple Line Plot of Percentage Spending vs. Market Size, colored by Media Type

To gain a clearer understanding of the general trend as opposed to the extremes, Figure 5 displays how media type expenditure varies by market size. As shown, larger markets tend to be composed of a higher percentage of digital marketing compared to smaller markets, which are dominated by radio advertisement.

IV. Conclusions

In conclusion, the analysis of advertising expenditure data from 2018 to 2023 reveals a dynamic advertising landscape shaped by the increasing influence of digital technology. The observed shift towards digital advertising, particularly evident in larger markets, underscores the growing preference for online platforms to engage with target audiences. This transition likely reflects broader societal trends, including the increase in internet usage and media consumption habits of consumers. Moreover, the manual selection model projected that the advertising budget will become 100% digital by 2033, highlighting the rapid pace of digital transformation within the industry.

Additionally, the analysis reveals distinct advertising strategies employed across different markets, emphasizing the significance of market-specific dynamics. Larger markets exhibit a clear preference for digital advertising, likely driven by robust digital infrastructure and widespread internet usage. In contrast, smaller markets tend to rely more heavily on traditional advertising avenues such as radio and newspapers. This disparity suggests that factors like digital accessibility and consumer behavior play a pivotal role in shaping advertising trends.

Both the stepwise AIC and manually selected models offer many insights into advertising spending, but each have their own limitations. The stepwise AIC model provides a comprehensive view of predictors but may miss nuanced relationships as indicated by the relatively low R-squared. On the other hand, the manually selected model overlooks complexity and assumes a linear relationship which may not fit the true underlying relationship. In future studies, researchers might overcome these limitations by integrating more extensive datasets covering extended timeframes and more predictors to gauge the continually changing Canadian advertising landscape.

V. References

"Broadcasting Current Trends – Canadian Agency Advertising Gross Spending – CMR (2018-2023)." Open Government Portal. open.canada.ca/data/en/dataset/c4fef546-d355-4a12-b995-dd8257f55430.