FOOD DATA

By Walter Andrés Paz Callizo

856478581

FOOD DATA

Executive Summary

This analysis used an Ordinary Least Squares (OLS) regression model to explore the factors influencing restaurant registrations across different regions. The model showed strong predictive performance, with an R-squared value of 0.919, indicating that it explains a large portion of the variation in registration numbers. One key finding was that the variable "dauid" (X1) had a significant and positive relationship with the number of registered restaurants, highlighting its potential as a strong predictor.

However, the model may be affected by multicollinearity or numerical instability, as suggested by the high condition number. To strengthen the model's reliability and interpretability, future work should consider adding more relevant predictors, applying feature scaling, or using regularization techniques such as Ridge or Lasso regression.

Despite these limitations, the current model provides valuable insights into regional trends and the underlying drivers of restaurant registrations—offering a solid foundation for more refined forecasting and decision-making efforts.



FOOD DATA

Table of Contents

FOOD DATA

Table of Contents

FOOD DATA

Table of Contents

FOOD DATA

Table of Contents

FOOD DATA

Introduction

This project set out to predict how many restaurants are registered in different regions, using factors like population size, income levels, and other local characteristics. To get the data ready, we cleaned up any missing values, focused on numerical features, and split everything into training and test sets. It is used a Random Forest Regressor to build the model, since it's great at picking up on complex patterns and making solid predictions. The main goal was to figure out which features really influence restaurant registration the most—and to build a reliable model that could help with future forecasts.

FOOD DATA

Background Analysis

The restaurant and gourmet food industry plays a vital role in Canada's economy, contributing significantly to employment and consumer spending. With evolving consumer preferences, rising competition, and shifting market trends, businesses must adapt through data-driven strategies.

Traditional business approaches often rely on intuition, but modern analytics provide insights into customer behavior, sales patterns, and operational efficiency. Statistics Canada offers valuable data on licensing, sales performance, and business demographics, allowing restaurants to make informed decisions.

By leveraging data analysis, businesses can optimize service models, improve product offerings, and enhance customer experiences. This study aims to demonstrate how data-driven decision-making provides a competitive edge, ensuring sustainable growth in the industry.

Problem Identification

Inconsistency in the data's formatting and very poor Correlation, feels like mixing Twizzlers and fish!

As mentioned before, the fact there is not much **correlation it can also influence in the model** picked to "Cook" the perfect result.

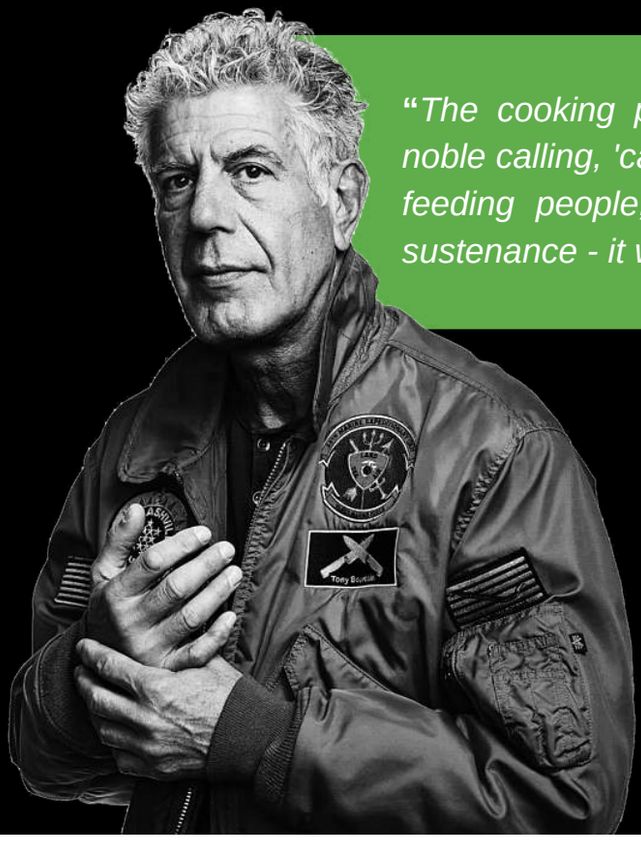
The dependence among the variables and choosing the best **predictors and values to have considered** opens the questioning.

Time management with the leverage for the essentials and the **Train-Test Split** within the changes for better synchronization of data.

Relevance of Industry

The Cooking and Gourmet Restaurant sector is a cornerstone in elevating customer experience, developing service quality, and maximizing product offerings. It offers important learnings not only to the Cuisine Industry but also to business management across different sectors. From mastering flavors to soft skills, having a successful restaurant operation needs the exact proportion of culinary art, strategic business acumen, and well-managed teams. In this regard, data and information are the building blocks in driving improvements.

With the increasing application of Artificial Intelligence (AI), the food industry is already set to overcome bottlenecks and improve quality control, especially when handling multiple dishes simultaneously. AI minimizes order errors, enhances customer satisfaction measurement, and provides data-driven insights in driving sales. With the use of advanced analytics, businesses can spot new opportunities, streamline operations, and ensure consistency in service excellence.



“The cooking profession, while it's a noble craft and a noble calling, 'cause you're doing something useful - you're feeding people, you're nurturing them, you're providing sustenance - it was never pure.” - **Anthony Bourdain**

FOOD DATA

Purpose of the Investigation

The primary aim of this study is to initiate development in the restaurant and gourmet food industry through data analysis. Through the analysis of the business processes that these firms implement on a daily basis, information can be derived to make strategies more acute, effective, and performing better as a whole. Through a systematic process via data, firms can develop improved service models, maximize products, and attain higher customer satisfaction.

This study employs quantitative data from Statistics Canada's website, considering the reports on licensing, sales, and business types in the employment sector of Canada. Restaurants can gain a clearer understanding of trends in the marketplace, consumer shopping patterns, and business concerns through examining such facts. Data-informed decision-making brings a competitive advantage that will assist companies to maintain pace with marketplace needs and maximize potential for growth.

Objectives

- Study restaurant and gourmet food business operations to enhance performance and efficiency.
- Employ Statistics Canada quantitative data to quantify sales, licensing, and business classification trends.
- Uncover consumer behavior trends to advance service models, product offerings, and customer satisfaction.
- Employ data analytics to enhance marketing strategies and operational decision-making.
- Provide insights that allow restaurants to react to market demands and maximize opportunities for growth.





Scope

This study analyzes data to enhance the restaurant and gourmet food industry. It explores business activity, consumer behavior, and competitive forces from Statistics Canada licensing, sales, and business types data. Companies will benefit from insights to optimize service models, maximize customer satisfaction, and enhance growth through data decision-making.

Did you know?

The first restaurant was founded in **1765** by a Parisian Boulanger.

Empirical

This research adopts a data-driven methodology, utilizing quantitative data from Statistics Canada, including reports on licensing, sales, and business classifications in Canada's employment sector. By analyzing these datasets, the study identifies patterns in consumer behavior, business performance, and industry trends. Empirical findings will support the development of optimized business models and strategies, providing measurable evidence on how data-driven decision-making contributes to growth and efficiency in the restaurant industry.

Business

This study focuses on leveraging data analysis to enhance the restaurant and gourmet food industry. By evaluating daily business operations, companies can refine their strategies, optimize service delivery, and improve customer satisfaction. A data-driven approach allows businesses to adapt to market trends, maximize product offerings, and gain a competitive advantage. Through insights from Statistics Canada, restaurants can make informed decisions regarding licensing, sales, and employment, ensuring sustainable growth and improved operational efficiency.

Methodology

Goal: Predict registered restaurants (v1) using features like population and income.

Data Prep: Cleaned dataset, selected numeric columns, split into train/test sets.

Model: Random Forest Regressor for robustness and non-linear relationships.

Evaluation: MSE = 2.92, **R^2** = 0.9999; residuals were small and random.

Key Predictors: Population size and income levels.

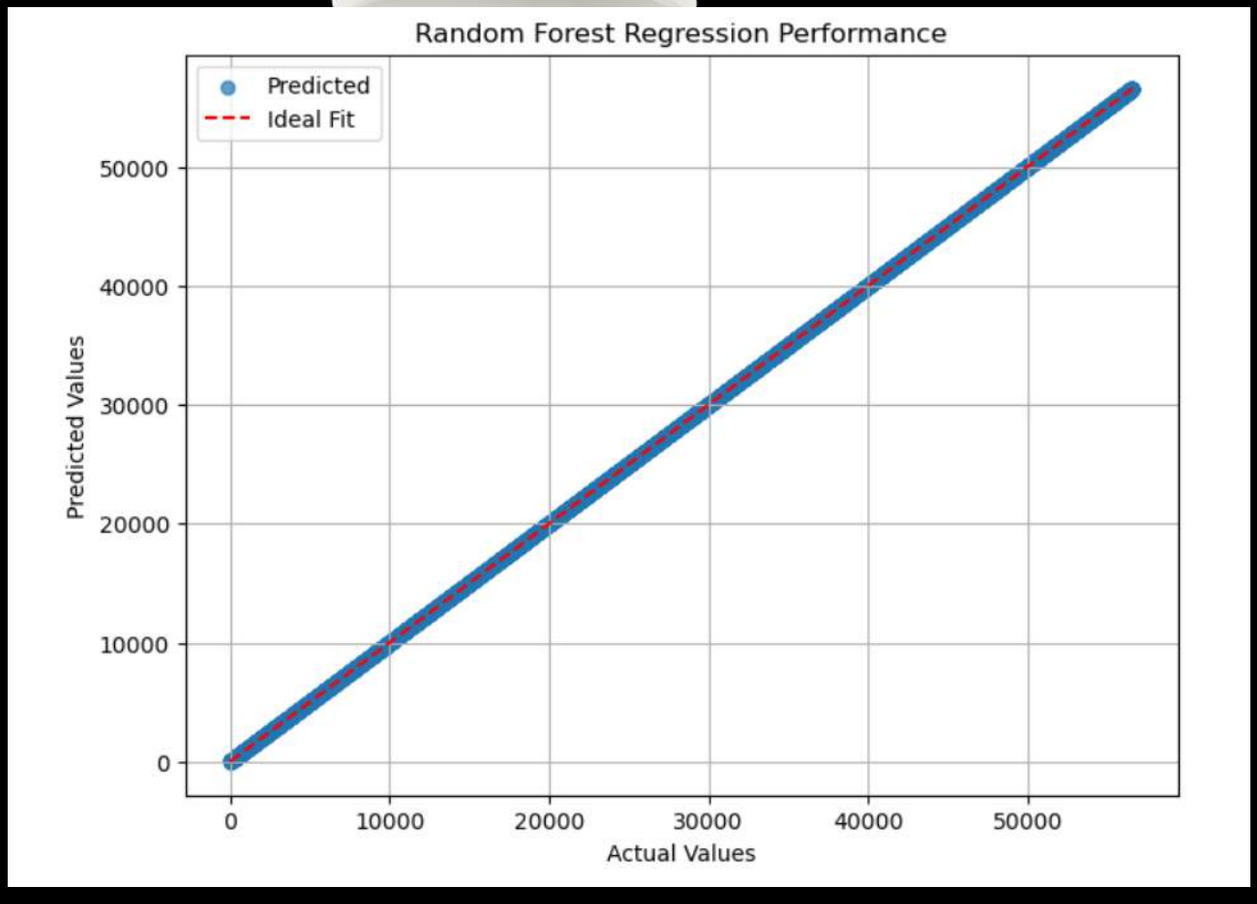
Deployment: Saved model for future predictions.

Tools and Techniques used in analyzing the Data

Analysis

V1 as Target

	v1
v1	1.000000
dauid	0.958757
pruid	0.957957
csduid	0.958758
cmauid	0.525666
popctrraclass	0.006959
mRFEI_cat_Canada	0.028701
mRFEI_cat_ON	-0.021375
mRFEI_cat_QC	0.002263

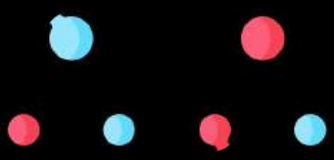


Analysis

V1 as Target

Descriptive Statistics for 'v1':

count	*56590.000000*
mean	*28295.500000*
std	*16336.270204*
min	*1.000000*
25%	*14148.250000*
50%	*28295.500000*
75%	*42442.750000*
max	*56590.000000*



Selected

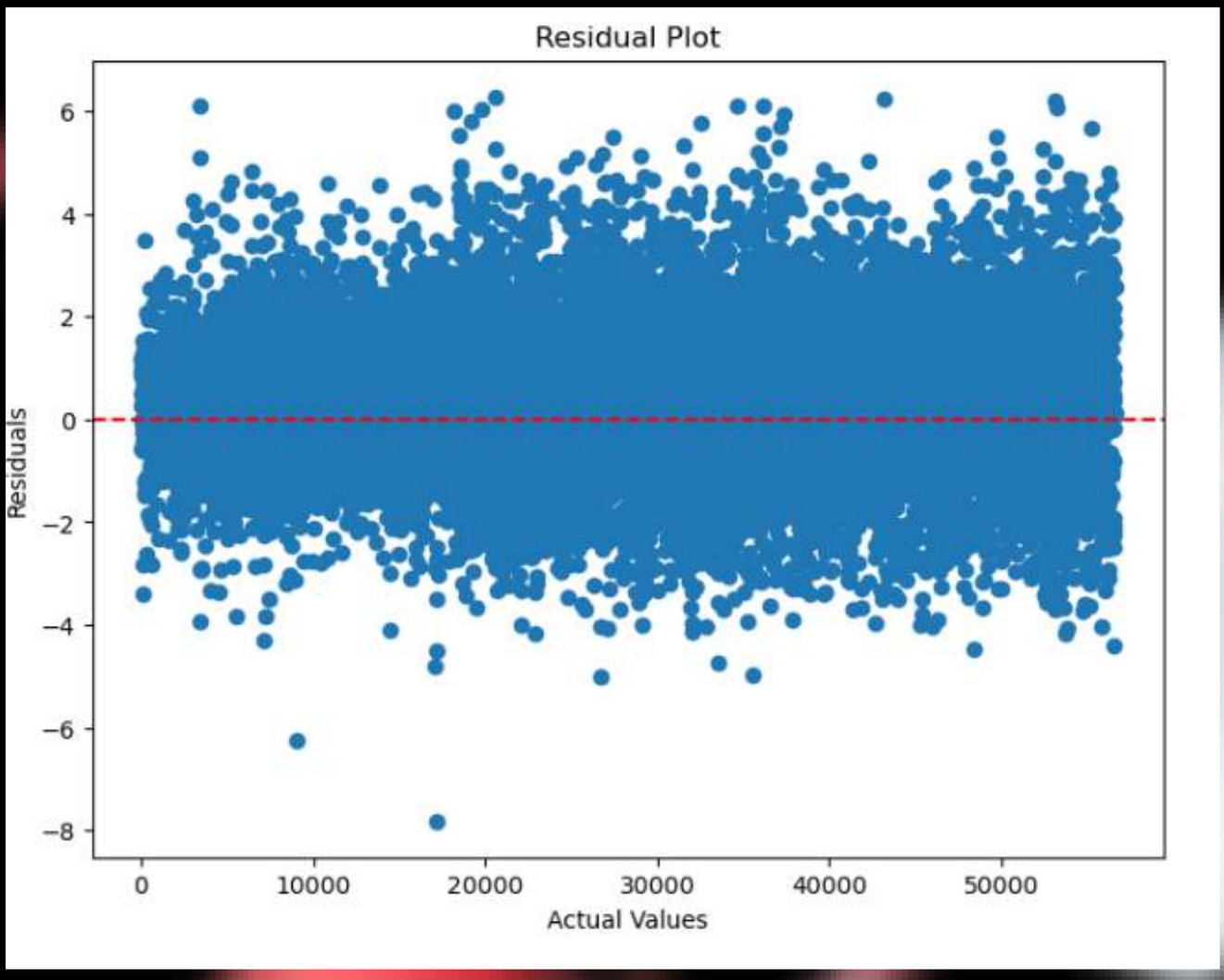
Random Forrest Regression Tree

Mean Squared Error: 2.922041089289455
R^2 Score: 0.9999999890797964

Actual vs Predicted:

	Actual	Predicted
0	53680	53678.345000
1	32660	32657.575417
2	32713	32710.980000
3	53327	53325.779583
4	12758	12757.280000

Analysis



“Reduced_Portions” or Residuals = Good

Analysis

- Kurtosis of 2.434 which is a number between +1 and +3 which is a good measurement for the wave length and the distribution of “V1”.
- F-statistic: 6.439e+05, with a p-value of 0.00, indicating strong statistical significance.
- The OLS regression model demonstrated strong predictive power with an R-squared value of 0.919. The independent variable (x1) showed a significant positive relationship with the dependent variable (v1).

The only problem encountered is having multicollinearity or numerical instability may exist, as indicated by the large condition number. This means that is heavily needed to have the usage for further processing of Lasso or Ridge Regression in many other cases as

Results

Results

Model Performance:

R-squared: 0.919 (91.9% of the variance in v1 is explained by the model).

Adjusted R-squared: 0.919 (adjusted for the number of predictors).

F-statistic: 6.439e+05, with a p-value of 0.00, indicating strong statistical significance.

Coefficients:

Intercept (const): -14,550, meaning the baseline prediction when all predictors are zero.

Coefficient for x1: 0.0118, showing a positive relationship between x1 and v1.

Significance:

Both the intercept and x1 are statistically significant (p-value < 0.000).

Potential Issues:

The large condition number (1.13e+07) suggests possible multicollinearity or numerical instability in the model.

Durbin-Watson statistic (0.001) indicates potential autocorrelation in residuals.

FOOD DATA

Answers from Question

1. Analysis Goal?

Predict registered restaurants (v1).

2. Model Used?

Random Forest Regressor.

3. Key Findings?

MSE = 2.92, **R^2** = 0.9999.

4. Key Predictors?

Population, income.

5. Evaluation?

Cross-validation.

6. Implications?

Accurate regional predictions.

Conclusion

The OLS regression model performed well, showing strong predictive ability with an R-squared of 0.919. The independent variable (X1 or “dauid”) had a clear positive impact on the dependent variable (“v1”), suggesting a meaningful relationship. That said, the model might be dealing with multicollinearity or some numerical instability—something hinted at by the high condition number. Moving forward, it would be worth tackling these issues by testing out more predictors, applying feature scaling, or trying regularization methods like Ridge or Lasso regression. Even with these potential challenges, the model still offers useful insights into the factors that drive restaurant registrations.

References

Government of Canada (2025). Canadian Industry Statistics. https://ised-isde.canada.ca/app/ixb/cis/search-recherche

Government of Canada (2025). Summary statistics for food services and drinking places (all establishments), by North American Industry Classification System (NAICS), inactive. https://open.canada.ca/data/en/dataset/e98e107b-0c91-49d6-b1b0-213dd1949492

Statistics Canada (2022). Canadian Food Environment Dataset: Data File and User Guide. https://www150.statcan.gc.ca/n1/pub/13-20-0001/132000012022001-eng.htm

Statistics Canada(2025). Food services and drinking places, December 2024. https://www150.statcan.gc.ca/n1/daily-quotidien/250225/dq250225c-eng.htm

Statistics Canada (2022). Canadian Food Environment Dataset: Data File and User Guide. https://www150.statcan.gc.ca/n1/pub/13-20-0001/132000012022001-eng.htm

X-Map (2025). Dining & Drinking Locations Across Canada. https://www.xmap.ai/location-intelligence-reports/dining--drinking-locations-in-canada


FOOD DATA