TTC Ridership Recovery and Changing Travel Patterns in Toronto

A Regime-Based Analysis of Pre-COVID, COVID/Recovery, and Post-COVID Ridership Dynamics

Project and Report by: Parsa Hassassedighi

Keywords: TTC ridership, regime-based analysis, regression modeling, Ridge and Lasso, dashboarding, Python, SAS

Source: https://www.reddit.com/r/TTC/comments/16bpr7b/shots_of_the_ttc_on_35mm_film/

Contents

1. Executive Summary

This project examines how TTC monthly average weekday ridership changed across **Pre-COVID**, **COVID/Recovery**, and **Post-COVID** periods, and how the factors associated with ridership shifted across those regimes.

The project was developed using **Python** for data preparation and dashboard prototyping, **SAS** for statistical modeling, and **Tableau** for visual analytics and dashboard design. Code and project files are available through the project GitHub repository, and the interactive dashboard is available through the deployed Render link.

A monthly analytical dataset was built by combining ridership, gas price, unemployment, and weather data. Exploratory analysis showed a clear structural break during COVID, which motivated a regime-based modeling approach. OLS regression, stepwise selection, Ridge, and Lasso were used to compare which variables mattered most in each period.

The results show that TTC ridership is not well described by one stable structure across the full timeline. **Pre-COVID ridership was dominated by annual seasonal persistence**, **COVID/Recovery ridership became more reactive to short-run conditions**, and **Post-COVID ridership showed a partial return to annual seasonal structure with some added unemployment and weather sensitivity**.

2. Business Questions and Ambitions

TTC ridership reflects more than transit usage alone. It is also related to commuting activity, labor-market conditions, travel cost, weather, and broader changes in mobility behavior. Because COVID disrupted all of these, historical ridership patterns could not simply be assumed to continue unchanged.

This project was guided by the following business questions:

1. How did TTC ridership change across pre-COVID, COVID/recovery, and post-COVID periods?
2. Which factors were most associated with ridership in each regime?
3. Did post-COVID ridership return to historical seasonal patterns?
4. Which variables gained or lost importance after COVID?

The ambition of the project was to go beyond descriptive visualization. The goal was to build a structured analytical dataset, apply regime-specific modeling, and communicate the findings clearly to both technical and non-technical audiences.

3. Project Scope and Analytical Framing

The project focuses on **Toronto TTC monthly average weekday ridership** and examines its relationship with selected external variables, mainly **gas prices**, **unemployment**, and **weather**. The final analysis is conducted at the **monthly level**, which is appropriate for medium-term trend and regime analysis, though not for highly detailed operational forecasting.

A regime-based framework was used because the pandemic introduced a major structural break in the ridership series. Pooling all time periods into one model would risk hiding meaningful changes in variable relationships. The timeline was therefore divided into three regimes:

- **Pre-COVID**
- **COVID/Recovery**
- **Post-COVID**

This project is designed as an explanatory and comparative analysis. It focuses on identifying changing relationships across time regimes rather than making strict causal claims.

4. Data Collection, Preparation, and Refinement

The final dataset was built by combining several ridership and contextual data sources into a single monthly analytical file. The core sources included:

- TTC monthly average weekday ridership
- monthly regular gas price data
- monthly unemployment rate data
- climate data, including temperature, precipitation, and snowfall
- derived time fields such as month, season, and COVID regime

The data preparation process involved several stages. Yearly climate files were appended into one climate dataset. Ridership data, which originally appeared in wide format, was reshaped into long monthly format. Date variables were standardized so that all sources could be merged consistently at the monthly level.

- Cleaning and refinement included:
- duplicate checking
- date conversion and standardization
- selecting relevant variables
- handling missingness
- aligning all datasets to a common time structure
- merging into one main monthly dataset

A deliberate decision was made **not to rely on broad imputation for missing climate values**, especially for precipitation and snowfall, because concentrated missingness in some periods could distort interpretation. Instead, the analysis proceeded using cleaned monthly weather variables with careful preprocessing.

Feature engineering was then used to create:

- "ridership_lag1"
- "ridership_lag12"
- "gas_price_lag1"
- "unemployment_rate_lag1"
- regime labels
- month dummy variables for regression models

The final modeling dataset contained **179 observations and 30 variables**, split into **109 pre-COVID**, **34 COVID/Recovery**, and **36 post-COVID** observations.

5. Exploratory and Statistical Analysis

5.1 Exploratory Analysis

Exploratory analysis was used first to understand the overall behavior of the ridership series before formal modeling. The main visual patterns were:

- a relatively stable and high ridership level in the pre-COVID period
- a sharp collapse during COVID
- a partial post-COVID recovery, without a full return to pre-COVID levels

Seasonality plots also showed that pre-COVID ridership followed a strong recurring monthly pattern, especially in the fall. This structure weakened considerably during the disruption period and then partially returned in the post-COVID period.

Initial relationship plots suggested that the link between ridership and external variables was not stable across the full series. This supported the decision to estimate separate models by regime instead of relying on one pooled specification.

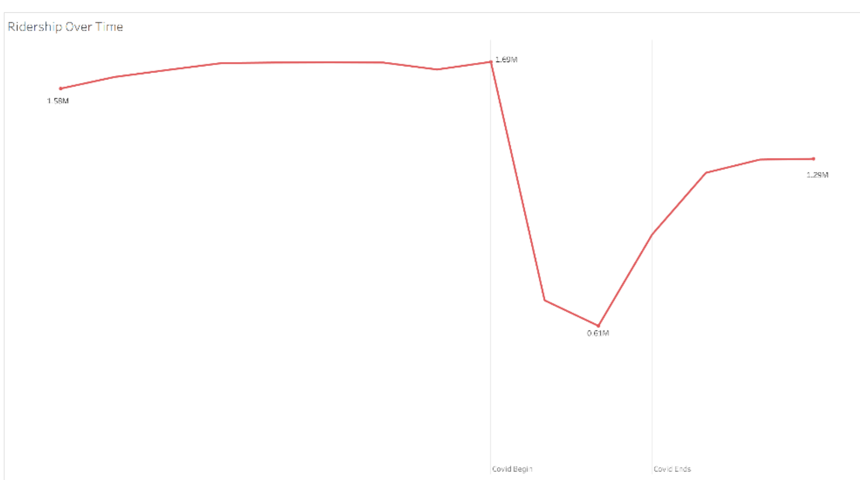


Figure 1: Ridership over Time

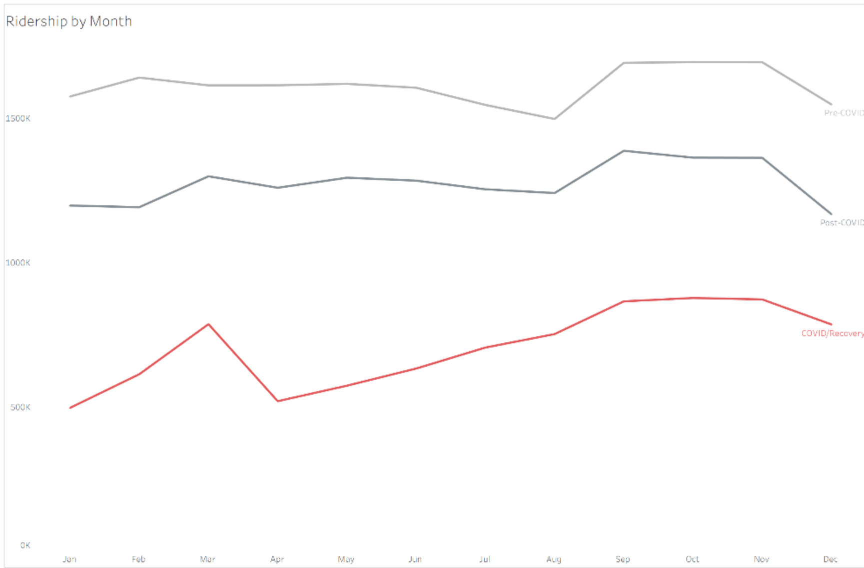


Figure 2: Ridership by Month and Regime

5.2 Statistical Methodology

The statistical phase used **multiple linear regression** as the primary framework because the outcome variable, monthly ridership, is continuous and the project objective was to compare explanatory structure across time periods.

Within each regime, four OLS model families were tested:

1. Month-only seasonal baseline
2. Lag memory model using "ridership_lag1" and "ridership_lag12"
3. Reduced external model adding lagged gas price and unemployment
4. External-only model without ridership history

Each model type served a different purpose:

- the seasonal baseline tested how much calendar structure alone explained ridership
- the lag memory model tested whether ridership mainly followed its own past
- the reduced external model tested whether economic variables added value beyond ridership memory
- the external-only model tested how much explanatory value existed without historical ridership structure

Stepwise regression was also used as a compact selection tool within each regime. It was not used as the only basis for interpretation, but it provided a concise summary of which variables remained in the model when weaker predictors were screened out.

Because some models contained overlapping predictors and regime-specific subsamples were relatively small, **Ridge and Lasso regression** were also added. Ridge was used to reduce coefficient instability under multicollinearity, while Lasso was used for variable selection by shrinking weak coefficients to zero. These regularized models were especially helpful for comparing variable importance across regimes.

5.3 Model Evaluation and Interpretation

The models were interpreted using:

- R^2 and Adjusted R^2
- Root MSE
- p-values
- coefficient signs and magnitudes
- VIF
- residual diagnostics
- Ridge/Lasso coefficient retention

The pre-COVID results are the most stable because that subsample is the largest. The COVID/Recovery and post-COVID analyses are still useful, but because those periods contain fewer observations, they are best interpreted as regime-comparison evidence rather than precise long-run structural inference.

6. Results and Interpretation

6.1 Pre-COVID Results

The pre-COVID period behaved as a **stable seasonal system**. The clearest explanatory pattern came from annual seasonal persistence rather than short-term fluctuation.

The month-only seasonal baseline explained **66.6%** of variation in pre-COVID ridership (**R^2 = 0.6661**), confirming the importance of recurring calendar structure. However, the lag memory model performed better (**R^2 = 0.7717**, Root MSE = 36,515), and within that model ridership_lag12 was highly significant while ridership_lag1 was not.

The reduced external model improved only slightly (**R^2 = 0.7867**, Root MSE = 35,631). In that model, ridership_lag12 remained the dominant variable, unemployment added modest value, and gas price was not significant. The external-only model was much weaker (**R^2 = 0.1887**), confirming that pre-COVID ridership was far better explained by historical ridership structure than by external conditions alone.

The regularized models reinforced this conclusion. Ridge placed the largest weight on ridership_lag12, and Lasso retained only ridership_lag12. Together, these results show that pre-COVID ridership was primarily governed by **annual seasonal repetition**, especially same-month prior-year demand.

6.2 COVID/Recovery Results

The COVID/Recovery period shows the strongest structural shift.

The month-only seasonal baseline performed poorly (**R^2 = 0.2274**, adjusted R^2 below zero), showing that the normal seasonal pattern had largely broken down. The lag memory model improved substantially (**R^2 = 0.6819**), and the reduced external model improved further to **R^2 = 0.8448**. In this regime, ridership_lag1 and gas_price_lag1 became much more important, while the pre-COVID dominance of ridership_lag12 weakened.

The external-only model also performed better here than in pre-COVID (**R^2 = 0.7067**), suggesting that external conditions carried more signal during the disruption period. Stepwise selection retained variables such as gas price and lagged ridership, and the regularized models told the same story: in Ridge, the strongest weights were on gas price and ridership_lag1, while in Lasso the retained variables were exactly **gas price and ridership_lag1**.

This indicates that during COVID/Recovery, TTC ridership became less of a stable annual system and more of a **short-run adaptive system**, reacting more strongly to recent conditions and disruption effects.

6.3 Post-COVID Results

The post-COVID period shows a **partial return to normal structure**, but not a full return to pre-COVID dynamics.

The month-only model explained **70.1%** of post-COVID variation (**R^2 = 0.7007**), indicating that seasonality had re-emerged. The lag memory model showed that ridership_lag12 was significant again while ridership_lag1 was not, meaning the system returned more toward annual seasonal structure than short-term momentum.

The reduced external model produced **R^2 = 0.6061**, with `ridership_lag12` significant and unemployment also significant, while gas price was not. In the external-only model, unemployment and snowfall remained significant, and precipitation was borderline. Stepwise selection retained only **ridership_lag12 and unemployment_rate_lag1**.

The regularized models align with this interpretation. Ridge again placed the largest weight on ridership_lag12, while Lasso retained ridership_lag12 along with snowfall and precipitation. These results suggest that post-COVID ridership has moved back toward annual seasonal memory, but with some added labor-market and weather sensitivity compared with pre-COVID.

6.4 Cross-Regime Comparison

The strongest conclusion of the project is that the importance of ridership drivers changed materially across the three regimes.

- **Pre-COVID:** annual seasonal persistence dominated
- **COVID/Recovery:** short-run dynamics and gas price became more important
- **Post-COVID:** annual memory returned, but with some added unemployment and weather influence

The system therefore moved from stable seasonal structure, to reactive disruption, and then to partial normalization. This regime-based interpretation is consistent across exploratory analysis, OLS models, stepwise results, and regularized models.

6.5 Dashboards and Visual Communication

Two dashboards were designed to communicate findings to different audiences:

- a **general dashboard**, focused on trend, recovery, and key business insights
- a **technical dashboard**, focused on driver comparison, coefficient patterns, and regime-specific variable importance

This split improved communication quality. The manager version emphasized the drop, recovery rate, and return of seasonality, while the technical version highlighted the shift from annual memory to short-run adaptation and back toward seasonal structure.

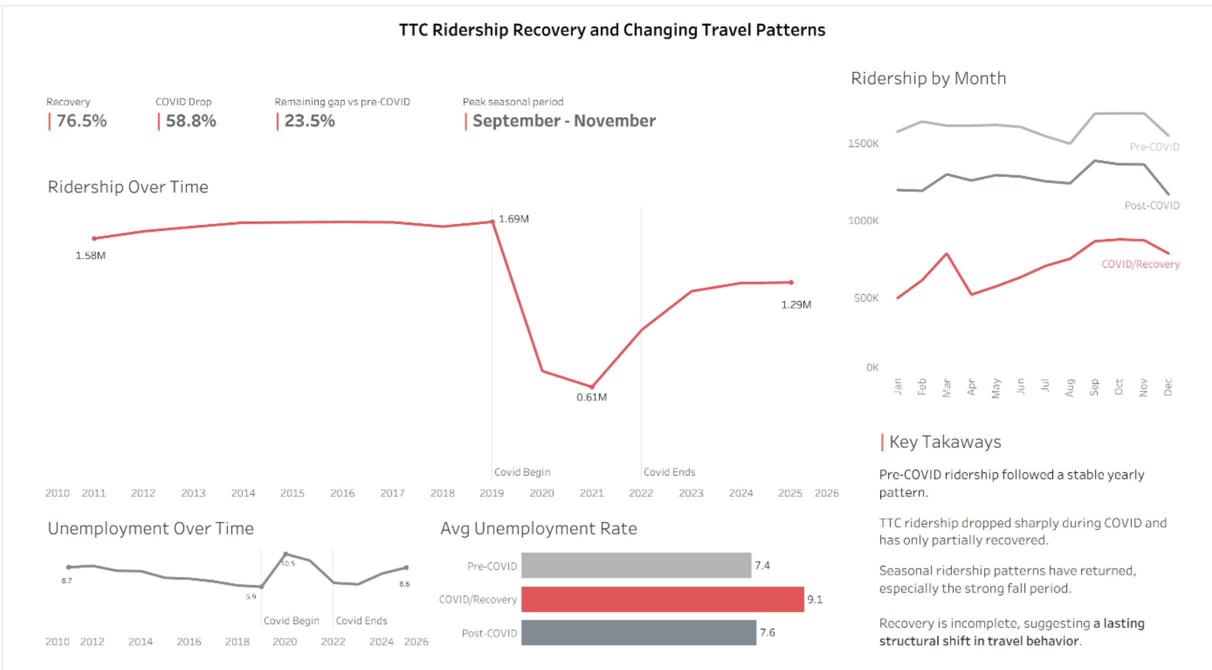


Figure 3: General Dashboard

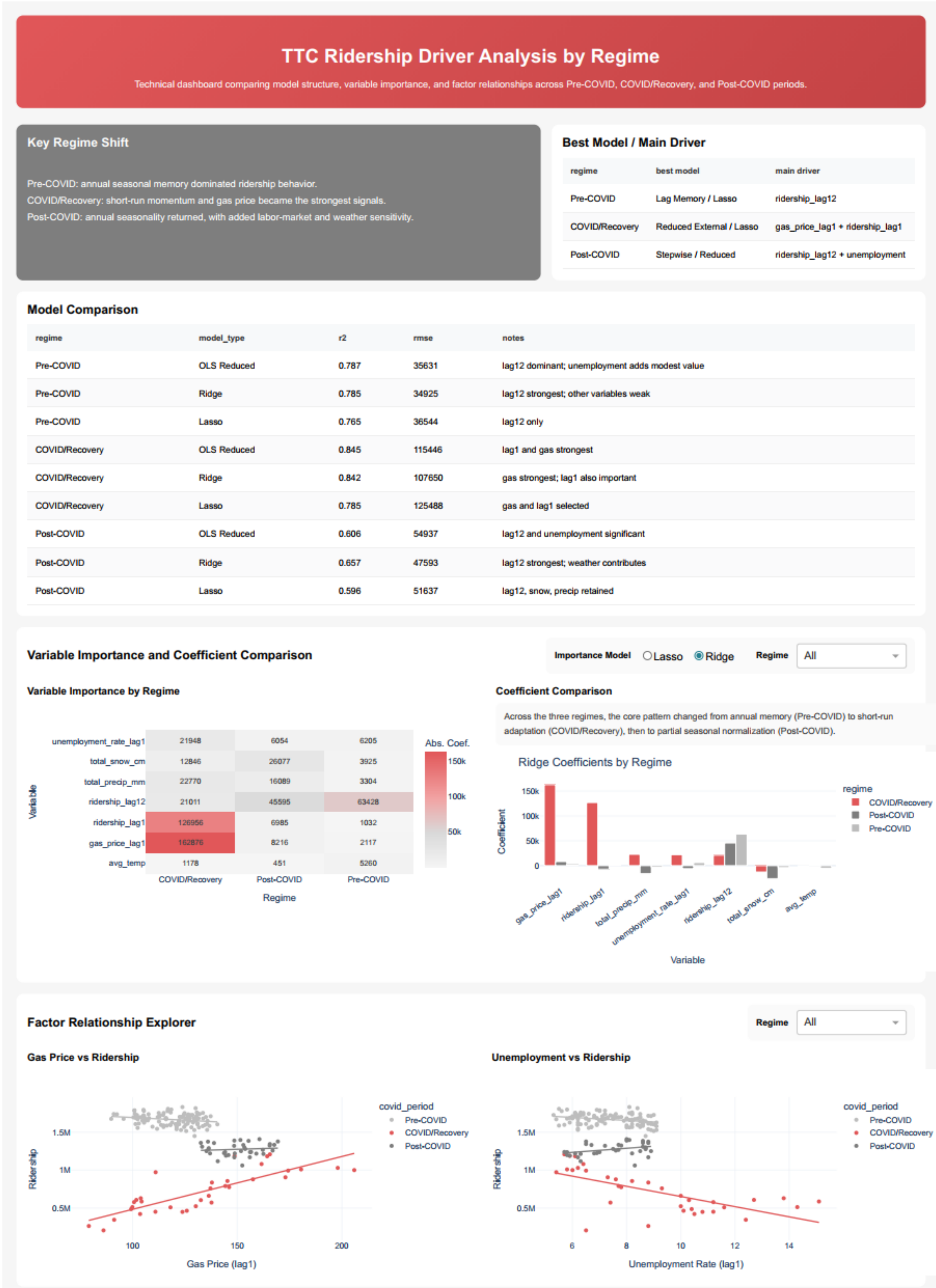


Figure 4: Technical Dashboard

Interactive version of the Technical Dashboard is available through the link bellow:

https://ttc-ridership-dashboard.onrender.com

7. Implications, Limitations, and Future Direction

The results have several practical implications. TTC ridership should not be modeled as one stable process across the entire timeline. Planning assumptions based only on pre-2020 patterns may not perform well when applied to disruption or recovery periods. The post-COVID period shows real recovery, but not a full return to the old demand structure.

There are also important limitations. The analysis is monthly, which is appropriate for strategic pattern analysis but not highly granular operational forecasting. The COVID/Recovery and post-COVID subsamples are relatively small, so those regime-specific estimates are less stable than the pre-COVID results. The models identify associations rather than prove causality. In addition, some potentially relevant behavioral variables, such as direct remote-work intensity, were not fully incorporated into the main model.

Future work could extend the project by:

- adding a remote work or office occupancy proxy
- extending the post-COVID sample as more data becomes available
- exploring route-level or mode-level ridership if accessible
- testing time-series forecasting models such as ARIMAX later
- comparing explanatory results with formal out-of-sample forecasting performance

8. Conclusion

This project shows that TTC ridership is best understood as a **regime-dependent system** rather than a single stable series. The pandemic changed not only ridership levels but also the relative importance of the factors associated with ridership.

Before COVID, ridership was primarily explained by annual seasonal persistence. During COVID/Recovery, ridership became more reactive to short-run conditions, especially recent ridership and gas price. In the post-COVID period, annual seasonal structure partially returned, but some labor-market and weather sensitivity remained.

The main takeaway is that the recovery of TTC ridership should be interpreted not only in terms of volume returning, but also in terms of **how the structure of demand changed across regimes**.

9. Appendices

Appendix A — Data Sources and Variable Dictionary

The final analytical dataset was constructed by combining monthly ridership data with economic and weather indicators relevant to TTC usage in Toronto. The main sources are listed below.

1. TTC Average Weekday Ridership

- Source: City of Toronto Dashboard
- Link: https://www.toronto.ca/city-government/data-research-maps/toronto-dashboard/?WT.rd_id=%2Fcity-government%2Fdata-research-maps%2Ftoronto-progress-portal%2F
- Frequency: Monthly

2. Daily Climate Data

- Source: Environment and Climate Change Canada
- Link: https://climate.weather.gc.ca/index_e.html
- Frequency: Daily, aggregated to monthly level for this project

3. Fuel Price Data

- Source: Statistics Canada, Table 18-10-0001-01
- Link: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1810000101
- Frequency: Monthly

4. Unemployment Rate

- Source: Statistics Canada, Table 14-10-0459-01
- Link: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1410045901
- Frequency: Monthly

Below is the main variable dictionary for the final modeling dataset used in the project.

Variable	Description	Type	Unit / Format
date	Monthly date field used as the primary time key	Date	YYYY-MM-DD
ridership	TTC average weekday ridership	Numeric	Riders
year	Calendar year extracted from date	Numeric	YYYY
month	Calendar month extracted from date	Numeric	1–12
covid_period	Regime label used to split the analysis into pre-COVID, COVID/Recovery, and post-COVID periods	Categorical	Text
gas_price_regular / gas_price_lag1	Monthly regular gasoline price, or its lagged version used in modeling	Numeric	Cents per litre
unemployment_rate / unemployment_rate_lag1	Monthly unemployment rate, or its lagged version used in modeling	Numeric	Percent
avg_temp	Monthly average temperature	Numeric	°C

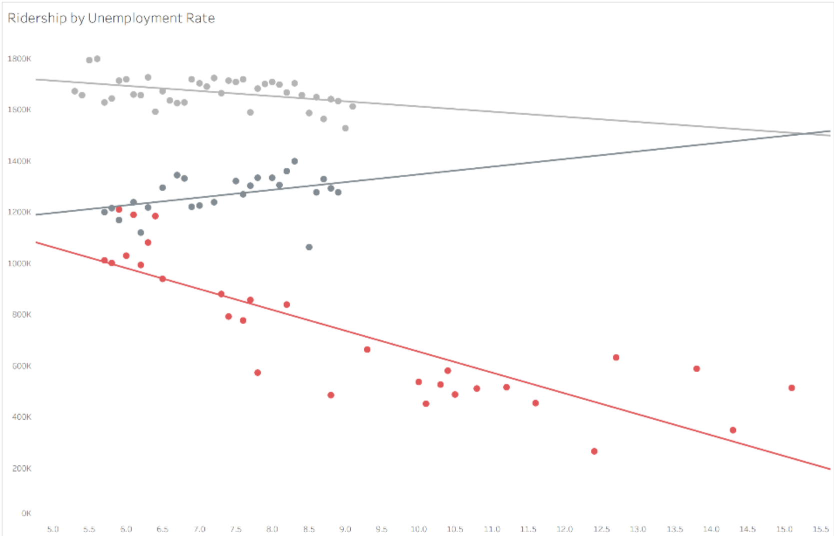
avg_max_temp	Monthly average maximum temperature	Numeric	°C
avg_min_temp	Monthly average minimum temperature	Numeric	°C
total_rain_mm	Monthly total rainfall	Numeric	mm
total_snow_cm	Monthly total snowfall	Numeric	cm
total_precip_mm	Monthly total precipitation	Numeric	mm
days_in_month	Number of days in each month	Numeric	Count
missing_temp_days	Number of missing daily temperature observations before aggregation	Numeric	Count
missing_precip_days	Number of missing daily precipitation observations before aggregation	Numeric	Count
missing_snow_days	Number of missing daily snowfall observations before aggregation	Numeric	Count
temp_completeness	Completeness indicator for monthly temperature data	Numeric	Ratio / Percent
precip_completeness	Completeness indicator for monthly precipitation data	Numeric	Ratio / Percent
snow_completeness	Completeness indicator for monthly snowfall data	Numeric	Ratio / Percent
climate_data_flag	Overall flag indicating climate data completeness / quality for the month	Categorical / Numeric	Project-defined
ridership_lag1	Ridership from previous month	Numeric	Riders
ridership_lag12	Ridership from same month in previous year	Numeric	Riders
gas_price_lag1	Gas price from previous month	Numeric	Cents per litre
unemployment_rate_lag1	Unemployment rate from previous month	Numeric	Percent
month_1 to month_11	Month dummy variables used in regression, with one month omitted as reference	Binary	0/1

Table 1: Variable Dictionary

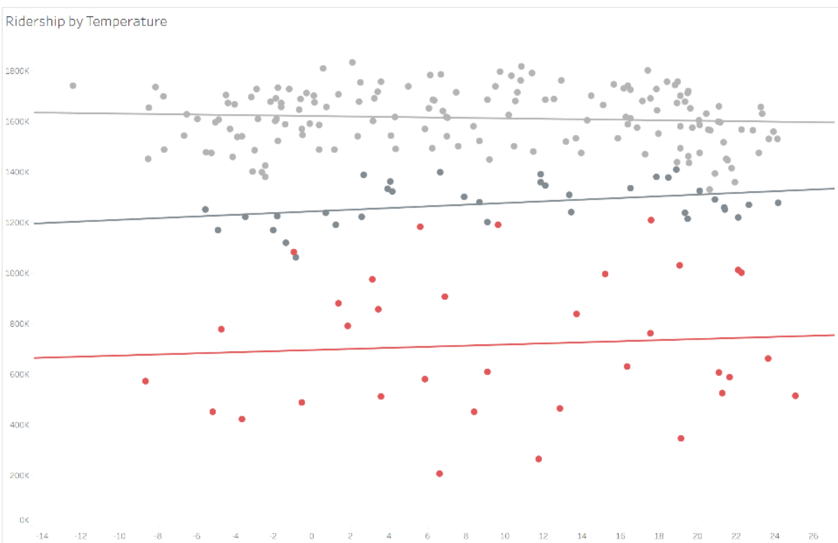
Note: Some variables were used mainly during cleaning or validation, while others were retained for statistical modeling and dashboard construction.

Appendix B — Additional Visuals

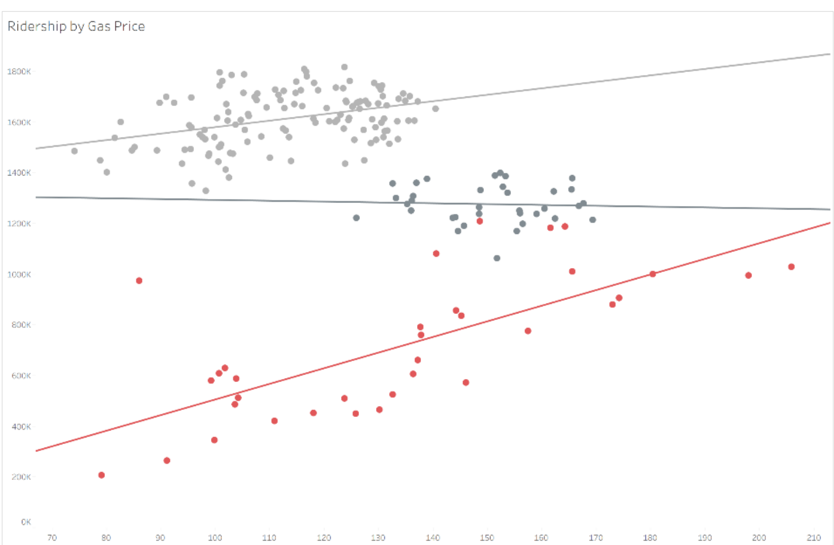
This appendix contains supplementary visuals that supported the exploratory and statistical analysis but were not included in the main body of the report. These figures provide additional context without overloading the main narrative.



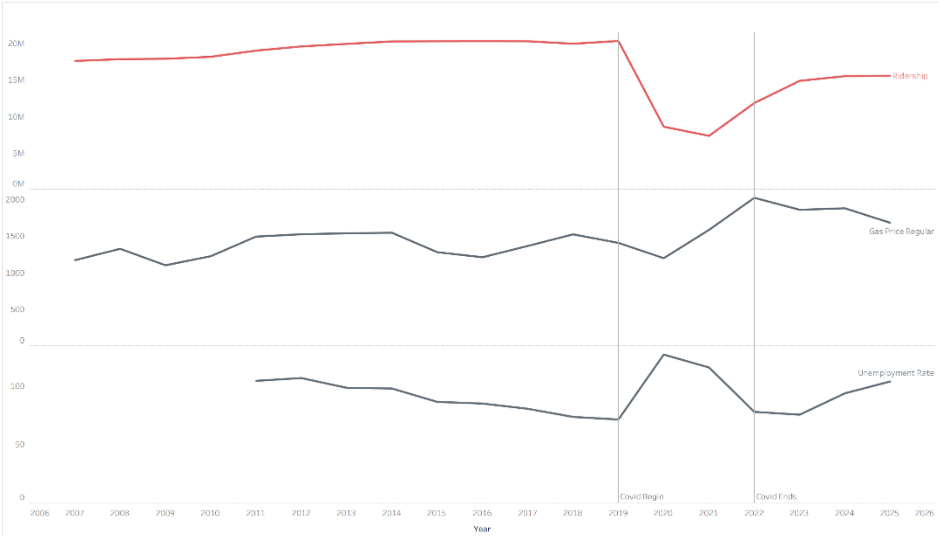
1-Ridership by Unemployment Rate



2-Ridership by Temperature



3-Ridership by Gas Price



4- SUM of Ridership, Gas Price, and Unemployment rate over time.

Appendix C — Full Regression Outputs

This appendix contains the full regression output tables used to support the model comparisons discussed in the main report. These materials are included here for completeness and technical transparency.

C.1 Pre-COVID OLS / PROC REG Outputs

Pre-COVID Model P1: Month-only Seasonal Baseline

The REG Procedure
Model: MODEL1
Dependent Variable: ridership

Number of Observations Read	109
Number of Observations Used	109

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	4.124335E11	37493954545	17.59	<.0001
Error	97	2.067225E11	2131159794		
Corrected Total	108	6.19156E11			

Root MSE	46164	R-Square	0.6661
Dependent Mean	1666000	Adj R-Sq	0.6283
Coeff Var	2.77098		

Program Summary - Ridership_Regression_timebased.sas

Parameter Estimates							
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Tolerance	Variance Inflation
Intercept	1	1594444	15388	103.62	<.0001	.	0
month_1	1	50444	21762	2.32	0.0226	0.54500	1.83486
month_2	1	98056	21211	4.62	<.0001	0.52153	1.91743
month_3	1	71333	21762	3.28	0.0015	0.54500	1.83486
month_4	1	76556	21762	3.52	0.0007	0.54500	1.83486
month_5	1	76556	21762	3.52	0.0007	0.54500	1.83486
month_6	1	69333	21762	3.19	0.0019	0.54500	1.83486
month_7	1	-222.22222	21762	-0.01	0.9919	0.54500	1.83486
month_8	1	-46889	21762	-2.15	0.0337	0.54500	1.83486
month_9	1	145778	21762	6.70	<.0001	0.54500	1.83486
month_10	1	157444	21762	7.23	<.0001	0.54500	1.83486
month_11	1	157333	21762	7.23	<.0001	0.54500	1.83486

Table 2: Pre-Covid Month-only Model

Pre-COVID Model P2: Lag Memory Model

The REG Procedure
Model: MODEL1
Dependent Variable: ridership

Number of Observations Read	109
Number of Observations Used	109

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	4.778188E11	2.389094E11	179.18	<.0001
Error	106	1.413372E11	1333369856		
Corrected Total	108	6.19156E11			

Root MSE	36515	R-Square	0.7717
Dependent Mean	1666000	Adj R-Sq	0.7674
Coeff Var	2.19180		

Parameter Estimates							
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Tolerance	Variance Inflation
Intercept	1	412307	86641	4.76	<.0001	.	0
ridership_lag1	1	-0.01155	0.04982	-0.23	0.8171	0.85112	1.17492

Program Summary - Ridership_Regression_timebased.sas

Parameter Estimates							
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Tolerance	Variance Inflation
ridership_lag12	1	0.77340	0.04406	17.55	<.0001	0.85112	1.17492

Table 3: Pre-Covid Lag Memory Model

Pre-COVID Model P3: Reduced External Model

The REG Procedure
Model: MODEL1
Dependent Variable: ridership

Number of Observations Read	109
Number of Observations Used	109

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	4.871199E11	1.2178E11	95.92	<.0001

Program Summary - Ridership_Regression_timebased.sas

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Error	104	1.320361E11	1269577551		
Corrected Total	108	6.19156E11			

Root MSE	35631	R-Square	0.7867
Dependent Mean	1666000	Adj R-Sq	0.7785
Coeff Var	2.13872		

Parameter Estimates							
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Tolerance	Variance Inflation
Intercept	1	208807	119721	1.74	0.0841	.	0
ridership_lag1	1	0.00988	0.04937	0.20	0.8418	0.82520	1.21183
ridership_lag12	1	0.82343	0.04698	17.53	<.0001	0.71282	1.40288
gas_price_lag1	1	141.85794	312.62610	0.45	0.6509	0.85704	1.16680
unemployment_rate_lag1	1	9271.45594	3681.06036	2.52	0.0133	0.75058	1.33231

Table 4: Pre-Covid Reduced External Model

Pre-Covid Model P4: Whithout Ridership History

The REG Procedure
Model: MODEL1
Dependent Variable: ridership

Number of Observations Read	109
Number of Observations Used	109

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	5	1.168612E11	23372245288	4.79	0.0006
Error	103	5.022948E11	4876648287		
Corrected Total	108	6.19156E11			

Root MSE	69833	**R-Square**	0.1887
Dependent Mean	1666000	**Adj R-Sq**	0.1494
Coeff Var	4.19166		

Parameter Estimates							
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**	**Tolerance**	**Variance Inflation**
Intercept	1	1919828	75600	25.39	<.0001	.	0
gas_price_lag1	1	-472.45569	643.37155	-0.73	0.4644	0.77730	1.28650
unemployment_rate_lag1	1	-19164	6648.63562	-2.88	0.0048	0.88377	1.13152
avg_temp	1	-3537.31988	1144.77565	-3.09	0.0026	0.35011	2.85627
total_precip_mm	1	-58.67445	194.28718	-0.30	0.7633	0.87204	1.14674
total_snow_cm	1	-1911.73713	652.47152	-2.93	0.0042	0.39241	2.54833

Table 5: Pre-Covid non-historical Model

C.2 Post-COVID OLS / PROC REG Outputs

COVID Model P1: Month-only Seasonal Baseline

The REG Procedure
Model: MODEL1
Dependent Variable: ridership

Number of Observations Read	36
Number of Observations Used	36

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	1.664229E11	15129353535	5.11	0.0004
Error	24	71096000000	2962333333		
Corrected Total	35	2.375189E11			

Root MSE	54427	R-Square	0.7007
Dependent Mean	1274556	Adj R-Sq	0.5635
Coeff Var	4.27030		

Parameter Estimates							
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Tolerance	Variance Inflation
Intercept	1	1167000	31424	37.14	<.0001	.	0
month_1	1	30000	44440	0.68	0.5061	0.54545	1.83333
month_2	1	24333	44440	0.55	0.5891	0.54545	1.83333
month_3	1	131333	44440	2.96	0.0069	0.54545	1.83333
month_4	1	91667	44440	2.06	0.0501	0.54545	1.83333
month_5	1	126333	44440	2.84	0.0090	0.54545	1.83333
month_6	1	116333	44440	2.62	0.0151	0.54545	1.83333
month_7	1	86333	44440	1.94	0.0639	0.54545	1.83333
month_8	1	73333	44440	1.65	0.1119	0.54545	1.83333
month_9	1	219667	44440	4.94	<.0001	0.54545	1.83333
month_10	1	196000	44440	4.41	0.0002	0.54545	1.83333
month_11	1	195333	44440	4.40	0.0002	0.54545	1.83333

Table 6: Post-Covid Month-only Model

COVID Model P2: Lag Memory Model

The REG Procedure
Model: MODEL1
Dependent Variable: ridership

Number of Observations Read	36
Number of Observations Used	36

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	1.195847E11	59792366115	16.73	<.0001
Error	33	1.179342E11	3573762323		
Corrected Total	35	2.375189E11			

Root MSE	59781	R-Square	0.5035
Dependent Mean	1274556	Adj R-Sq	0.4734
Coeff Var	4.69034		

Parameter Estimates							
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Tolerance	Variance Inflation
Intercept	1	991495	148492	6.68	<.0001	.	0
ridership_lag1	1	-0.09302	0.14218	-0.65	0.5175	0.65033	1.53769
ridership_lag12	1	0.34258	0.06822	5.02	<.0001	0.65033	1.53769

Table 7: Post-Covid Lag Memory Model

COVID Model P3: Reduced External Model

The REG Procedure
Model: MODEL1
Dependent Variable: ridership

Number of Observations Read	36
Number of Observations Used	36

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	1.439579E11	35989487302	11.92	<.0001
Error	31	93560939683	3018094828		
Corrected Total	35	2.375189E11			

Root MSE	54937	R-Square	0.6061
Dependent Mean	1274556	Adj R-Sq	0.5553
Coeff Var	4.31030		

Parameter Estimates							
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Tolerance	Variance Inflation
Intercept	1	1150372	229458	5.01	<.0001	.	0
ridership_lag1	1	-0.12783	0.13135	-0.97	0.3380	0.64348	1.55405
ridership_lag12	1	0.53766	0.09792	5.49	<.0001	0.26656	3.75151
gas_price_lag1	1	-231.08049	1009.47814	-0.23	0.8204	0.73671	1.35738
unemployment_rate_lag1	1	-41068	16097	-2.55	0.0159	0.29056	3.44159

Table 8: Post-Covid Reduced External Model

Covid Model P4: Whithout Ridership History

The REG Procedure
Model: MODEL1
Dependent Variable: ridership

Number of Observations Read	36
Number of Observations Used	36

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	1.24775E11	24954996154	6.64	0.0003
Error	30	1.127439E11	3758130271		
Corrected Total	35	2.375189E11			

Root MSE	61304	R-Square	0.5253
Dependent Mean	1274556	Adj R-Sq	0.4462
Coeff Var	4.80980		

Parameter Estimates							
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Tolerance	Variance Inflation
Intercept	1	882329	203872	4.33	0.0002	.	0
gas_price_lag1	1	1852.73421	1098.41748	1.69	0.1020	0.77481	1.29064
unemployment_rate_lag1	1	25816	10814	2.39	0.0235	0.80159	1.24752
avg_temp	1	-582.58563	1682.64491	-0.35	0.7316	0.43838	2.28112
total_precip_mm	1	-653.05415	321.34629	-2.03	0.0511	0.83161	1.20249
total_snow_cm	1	-3248.84265	1045.26520	-3.11	0.0041	0.46939	2.13040

Table 9: Post-Covid non-historical Model

C.3 COVID/Recovery OLS / PROC REG Outputs

COVID Model P1: Month-only Seasonal Baseline

The REG Procedure
Model: MODEL1
Dependent Variable: ridership

Number of Observations Read	34
Number of Observations Used	34

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	5.662673E11	51478845811	0.59	0.8178
Error	22	1.923675E12	87439780303		
Corrected Total	33	2.489942E12			

Root MSE	295702	R-Square	0.2274
Dependent Mean	715471	Adj R-Sq	-0.1589
Coeff Var	41.32975		

Parameter Estimates							
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Tolerance	Variance Inflation
Intercept	1	785333	170724	4.60	0.0001	.	0
month_1	1	-288833	269938	-1.07	0.2962	0.63750	1.56863
month_2	1	-172333	269938	-0.64	0.5298	0.63750	1.56863
month_3	1	1000.00000	241440	0.00	0.9967	0.54839	1.82353
month_4	1	-265667	241440	-1.10	0.2831	0.54839	1.82353
month_5	1	-212000	241440	-0.88	0.3894	0.54839	1.82353
month_6	1	-152667	241440	-0.63	0.5337	0.54839	1.82353
month_7	1	-80000	241440	-0.33	0.7435	0.54839	1.82353
month_8	1	-33333	241440	-0.14	0.8914	0.54839	1.82353
month_9	1	80000	241440	0.33	0.7435	0.54839	1.82353
month_10	1	91667	241440	0.38	0.7078	0.54839	1.82353
month_11	1	86667	241440	0.36	0.7230	0.54839	1.82353

Table 10: Covid/Recovery Month-only Model

COVID Model P2: Lag Memory Model

The REG Procedure
Model: MODEL1
Dependent Variable: ridership

Number of Observations Read	34
Number of Observations Used	34

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	1.697893E12	8.489466E11	33.23	<.0001
Error	31	7.920493E11	25549978977		
Corrected Total	33	2.489942E12			

Root MSE	159844	R-Square	0.6819
Dependent Mean	715471	Adj R-Sq	0.6614
Coeff Var	22.34105		

Parameter Estimates							
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Tolerance	Variance Inflation
Intercept	1	408785	91887	4.45	0.0001	.	0
ridership_lag1	1	0.61225	0.08865	6.91	<.0001	0.96595	1.03526
ridership_lag12	1	-0.14773	0.04955	-2.98	0.0055	0.96595	1.03526

Table 11: Covid/Recovery Lag Memory Model

COVID Model P3: Reduced External Model

The REG Procedure
Model: MODEL1
Dependent Variable: ridership

Number of Observations Read	34
Number of Observations Used	34

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	2.103436E12	5.258589E11	39.46	<.0001
Error	29	3.86507E11	13327826100		
Corrected Total	33	2.489942E12			

Root MSE	115446	R-Square	0.8448
Dependent Mean	715471	Adj R-Sq	0.8234
Coeff Var	16.13570		

Parameter Estimates							
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Tolerance	Variance Inflation
Intercept	1	-738581	254759	-2.90	0.0071	.	0
ridership_lag1	1	0.48653	0.11177	4.35	0.0002	0.31696	3.15496
ridership_lag12	1	0.04711	0.05814	0.81	0.4244	0.36592	2.73287
gas_price_lag1	1	6298.35822	1158.95146	5.43	<.0001	0.28929	3.45677
unemployment_rate_lag1	1	23274	14003	1.66	0.1073	0.28107	3.55788

Table 12: Covid/Recovery Reduced External Model

Covid Model P4: Whithout Ridership History

The REG Procedure
Model: MODEL1
Dependent Variable: ridership

Number of Observations Read	34
Number of Observations Used	34

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	1.75961E12	3.51922E11	13.49	<.0001
Error	28	7.303326E11	26083306505		
Corrected Total	33	2.489942E12			

Root MSE	161503	R-Square	0.7067
Dependent Mean	715471	Adj R-Sq	0.6543
Coeff Var	22.57301		

Parameter Estimates							
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Tolerance	Variance Inflation
Intercept	1	119369	278462	0.43	0.6714	.	0
gas_price_lag1	1	5725.07417	1334.52782	4.29	0.0002	0.42698	2.34203
unemployment_rate_lag1	1	-17049	16047	-1.06	0.2971	0.41881	2.38770
avg_temp	1	-3641.14481	5148.52398	-0.71	0.4853	0.32623	3.06532
total_precip_mm	1	941.64677	1055.83406	0.89	0.3801	0.85746	1.16623
total_snow_cm	1	-3472.08099	3460.59856	-1.00	0.3243	0.39073	2.55928

Table 13: Covid/Recovery non-historical Model

C.4 Supporting Diagnostic Outputs

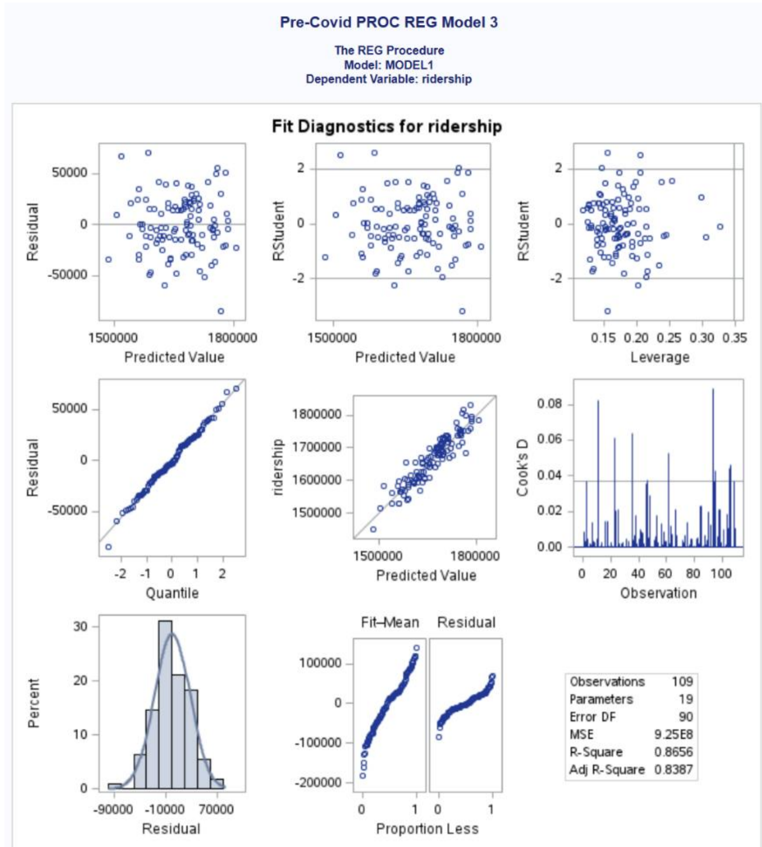


Table 14: Pre-Covid Residual Diagnostics Plots

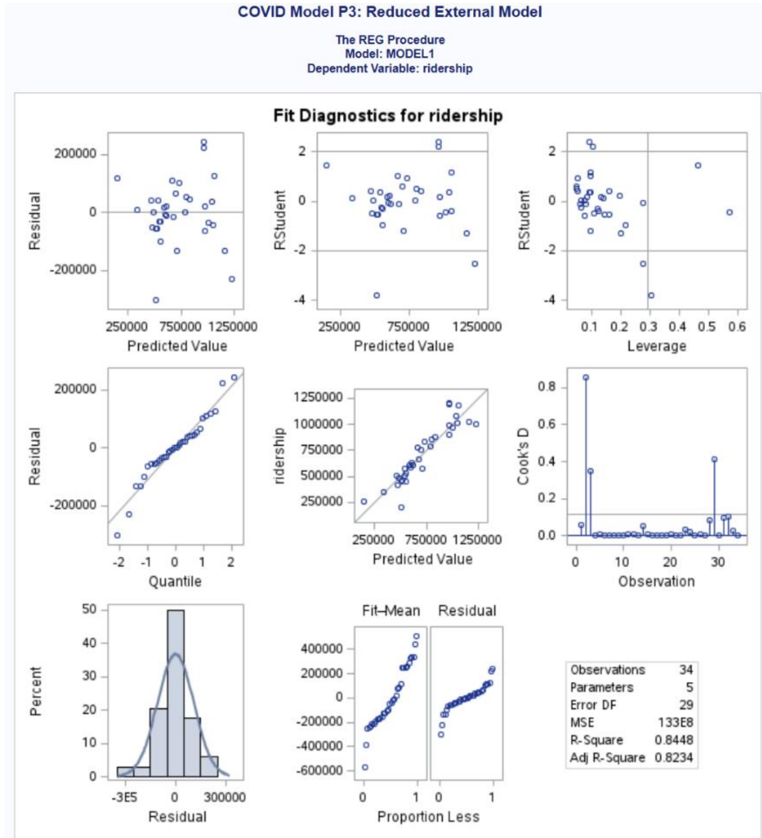


Table 15: Covid/Recovery Residual Diagnostics Plots

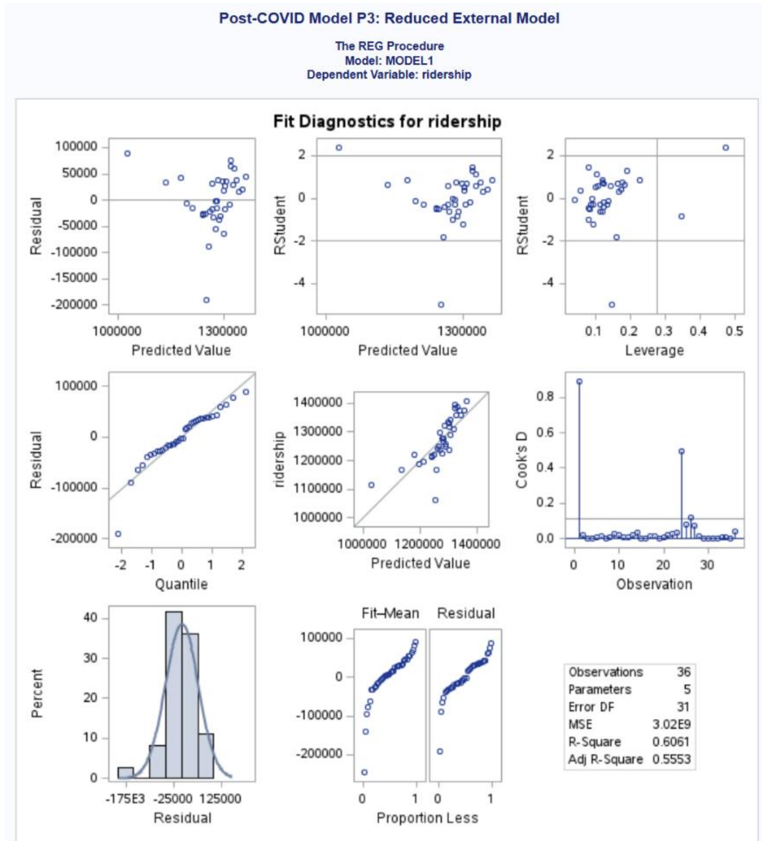


Table 16: Post-Covid Residual Diagnostics Plots

Appendix D — Ridge and Lasso Outputs

This appendix contains the regularized regression outputs used to compare variable importance under shrinkage and selection across the three regimes.

D.1 Ridge Regression

	Pre-COVID	COVID/Recovery	Post-COVID
ridership_lag12	63427.73958	21010.83615	45595.04393
unemployment_rate_lag1	6204.84817	21947.65881	-6054.024858
avg_temp	-5260.267676	-1177.916094	450.5146398
total_snow_cm	-3925.341452	-12845.54964	-26077.32694
total_precip_mm	-3303.762832	22770.21249	-16088.94108
gas_price_lag1	2117.372991	162875.5735	8216.475115
ridership_lag1	1032.012586	126956.1107	-6984.933879

Table 17: Ridge Regression Summary

D.2 Lasso Regression

	Pre-COVID	COVID/Recovery	Post-COVID
ridership_lag12	60030.00974	0	38127.11269
ridership_lag1	0	96169.20921	0
gas_price_lag1	0	124976.1705	319.7784844
unemployment_rate_lag1	0	0	0
avg_temp	0	0	0
total_precip_mm	0	0	-6307.56588
total_snow_cm	0	0	-22051.97231

Table 18: Lasso Regression Summary