Big Data Energy Solutions

x

Itty-Bitty Electric Company

Presented By:

Brandon Tang, Amelie Carrillo, Gerardo Gutierrez, Oscar Ramirez, Mark McLaughlin

AGENDA

1. EXECUTIVE OVERVIEW
2. DATA ANALYSIS PROCESS
3. RESEARCH & INSIGHTS
4. RECOMMENDATIONS & STRATEGY
5. IMPLEMENTATION PLAN
6. CONCLUSION

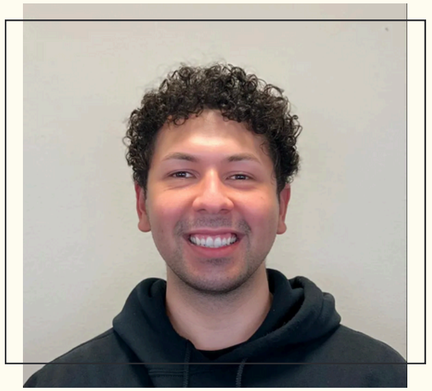
MEET THE TEAM

Founder & CEO

Brandon Tang

Research & Strategy Analyst

Amelie Carrillo



Research & Insights Analyst

Gerardo Gutierrez

Technical Lead & ML Engineer

Oscar Ramirez

Quality Analyst & Writer

Mark McLaughlin

EXECUTIVE OVERVIEW

BACKGROUND

IBEC is facing growing challenges in maintaining **grid reliability**, driven by **aging infrastructure, increasing environmental risks, and limited operational resources**. Frequent and prolonged outages have caused *Customer Minutes of Interruption (CMI)* to remain high across multiple service regions.

Without a targeted, data-driven strategy, IBEC risks **inefficient infrastructure investments, rising operational costs, and declining customer satisfaction**. To strengthen long-term system reliability, a clear analytical approach is needed to **identify** high-priority circuits, **address** root causes of outages, and **optimize** resource allocation.

SWOT ANALYSIS

Strengths

- Access to detailed **internal** outage data across **multiple** regions and circuits
- Established performance metrics to guide infrastructure **investment** decisions
- Strong organizational commitment to improving grid **reliability** and customer **satisfaction**

Weaknesses

- No **specific** geographic/location data in relation to Southern California
- Internal data **differs significantly** from industry, making **benchmarking** difficult
- Only around 100 records, **limiting** trend detection, forecasting, and **statistical confidence**

Opportunities

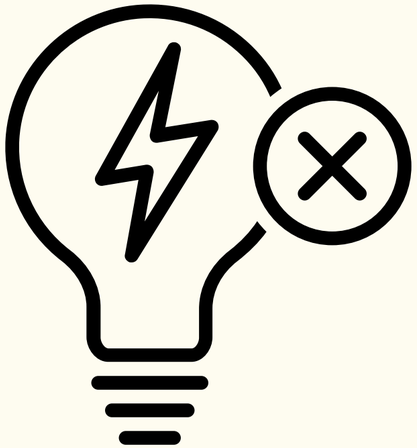
- Use **existing** CMI, SAIDI, and SAIFI data to rank and **prioritize** circuits
- Establish **foundation** for internal dashboards to track outage causes
- Demonstrate the **value** of analytics **internally**

Threats

- Risk of **misallocating** capital investments
- Potential reputational **damage** and **decline** of customer return rate
- Growing environmental or **extreme** weather events

OUR GOAL

Reduce outages, enhance system **reliability**, and maximize IBEC's infrastructure **investments** through smarter, data-driven **decisions**.



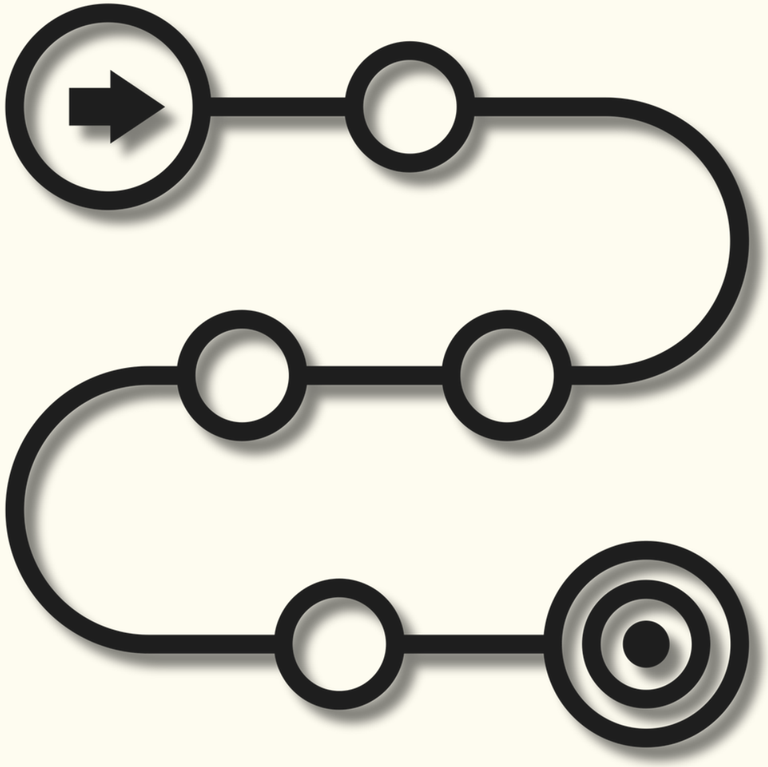
OBJECTIVES

- Identify outage causes that contribute the most to **Customer Minutes Interrupted (CMI)**
- Rank circuits by **impact** and **prioritize** for intervention
- Create dashboards and graphs to **visualize** trends
- Develop data-driven recommendations to **reduce** outages and **improve** reliability

SCOPE

- Focused on **circuit-level** outage data provided by IBEC
- Evaluated **key metrics**:
  - CMI
  - Outage Duration
  - Customers Affected
  - SAIDI
  - SAIFI

OUR PROCESS



DATA ANALYSIS PROCESS

Identify
#1

Define key objectives by pinpointing critical outage metrics and *high-impact* circuits.

Hypothesize
#2

Develop initial assumptions and analytical questions to guide *focused data exploration*.

Clean & Enrich
#3

Prepare the dataset by *addressing* missing values, inconsistencies, and *enhancing* data quality.

Visualize
#4

Create clear, *targeted* visuals to uncover trends, support insights, and communicate findings *effectively*.

Apply
#5

Translate analytical insights into *actionable* recommendations to drive *strategic* improvements for IBEC.

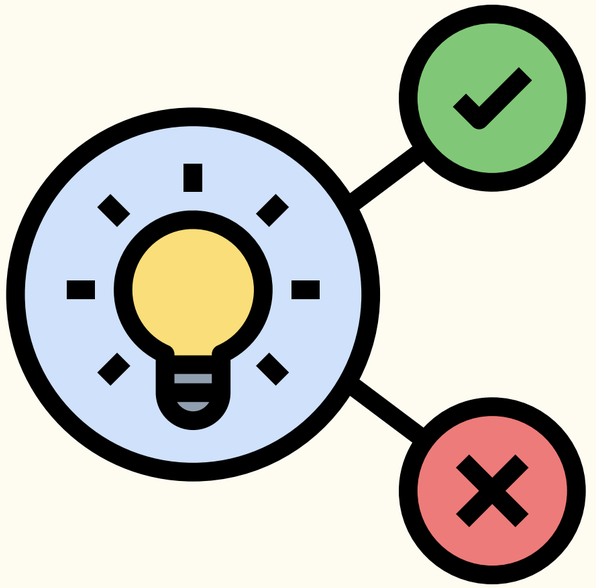
IDENTIFY

Before diving into the analysis, it's essential to first establish the key metrics and priorities that will guide our work. Our initial focus includes:

- Prioritizing circuits with high CMI
- Focusing on key outage patterns

HYPOTHESIZE

With key priorities identified, we developed initial hypotheses to guide our exploration and uncover root causes of outages.



Examples:

- *Which circuits have the highest CMI?*
- *What are the most frequent outage causes, and where do they occur?*
- *Are there seasonal patterns that could inform targeted maintenance?*

CLEAN & ENRICH

Enrichment Features

- *2024 SAIDI & SAIFI*
- *SAIDI & SAIFI Change from Previous Year*
- *Total Outage Time*
- *Total Outages*
- *Season and Month*

VISUALIZE

Choosing Suitable Visualizations:

- Focused on **categorical data** such as outage causes and regions
- *Bar charts* and *line graphs* were most **effective** for highlighting trends and comparisons

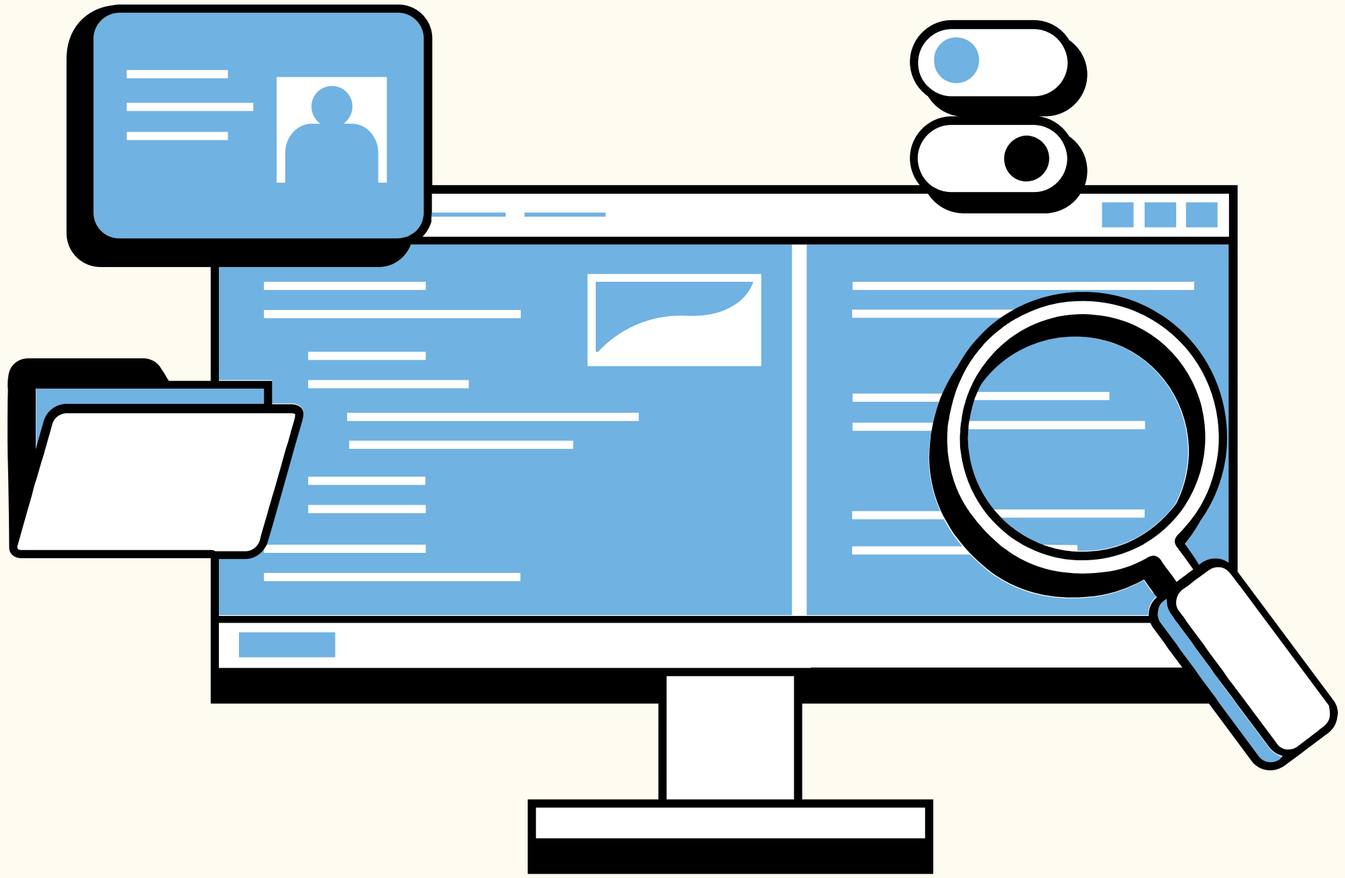
Tools & Methods Used:

- Utilized Python's libraries like *Pandas*, *Plotly Express*, and *Seaborn*
- Leveraged Tableau dashboards to create **dynamic**, **interactive insights** for easier decision-making

ANALYSIS & APPLICATION

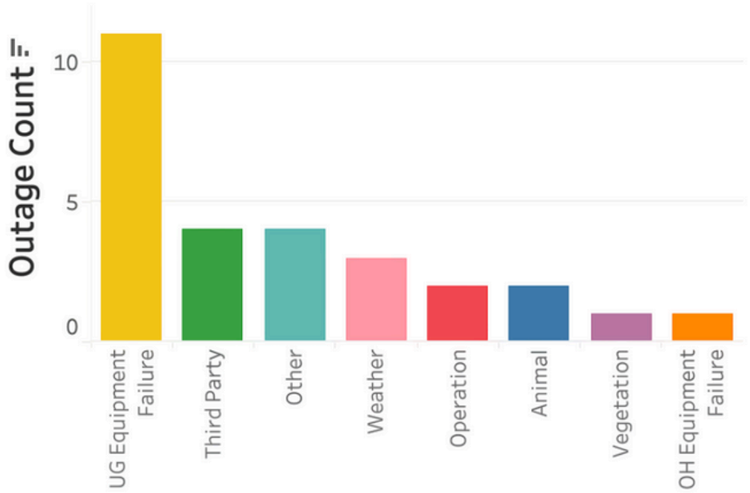
Performed a detailed analysis of outage patterns, applying performance metrics to ***prioritize*** *circuit upgrades and translate insights into* ***actionable*** *recommendations.*

RESEARCH & INSIGHTS

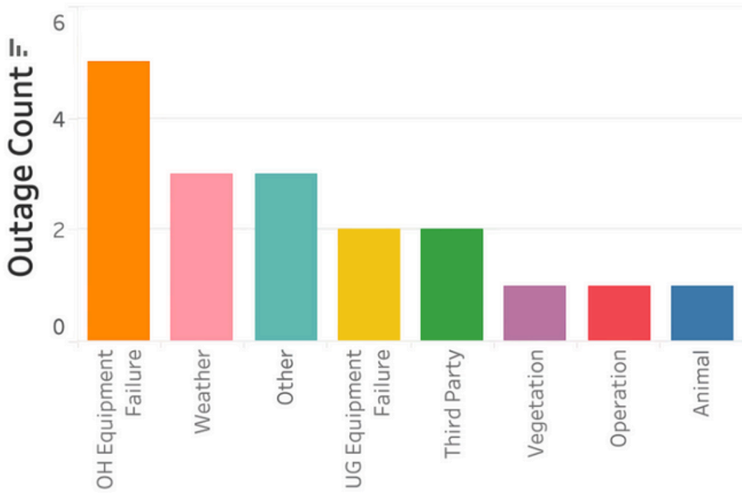


Outage Causes by Region - Dashboard

Coastal Region



Desert Region

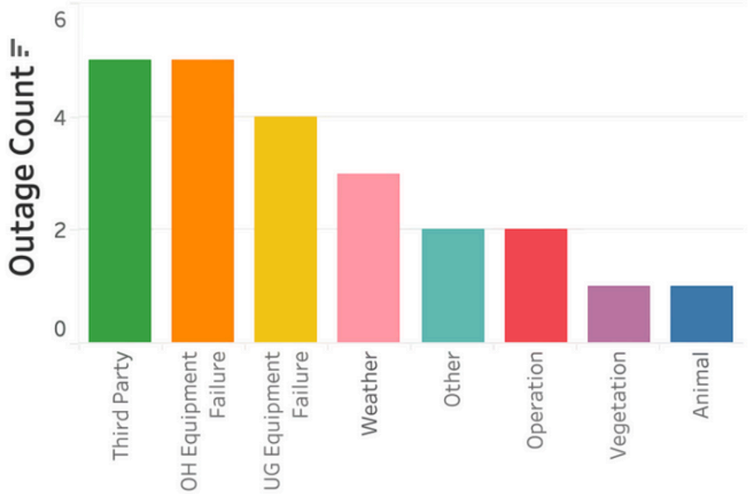


Outage Cause

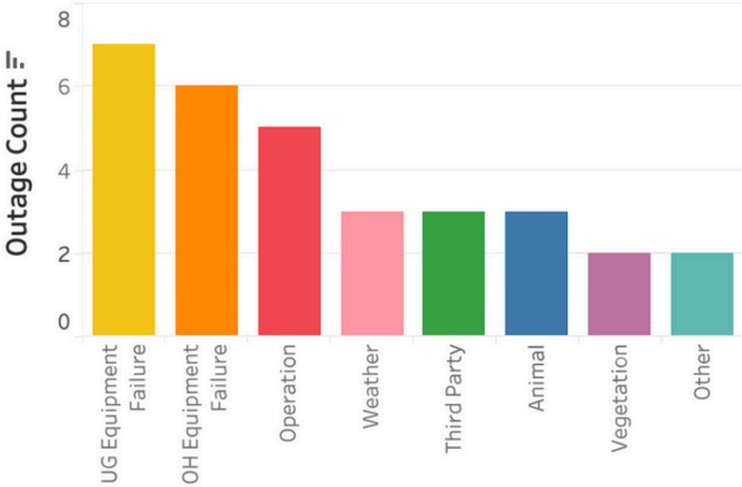
- Animal
- OH Equipment Failure
- Operation
- Other
- Third Party
- UG Equipment Failure
- Vegetation
- Weather

Region	Total CMI	Outage Count
Coastal	21,451,249	28
Desert	7,696,578	18
Mountain	21,224,894	23
North	25,597,270	31

Mountain Region

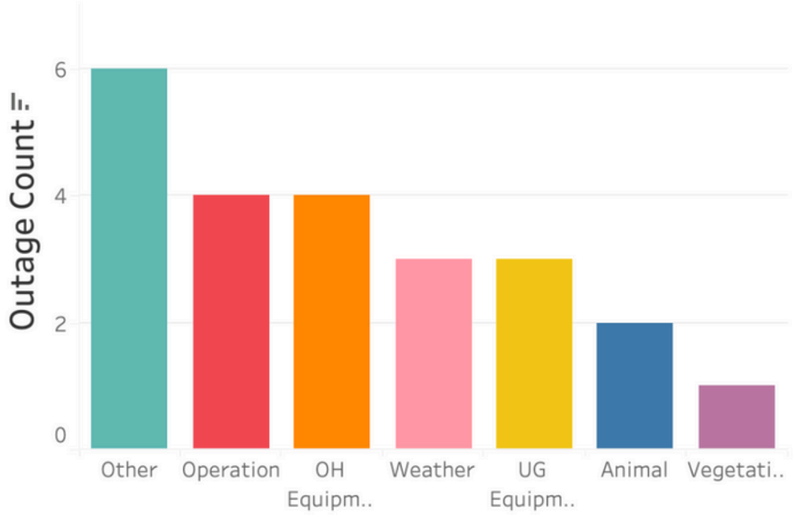


Northern Region

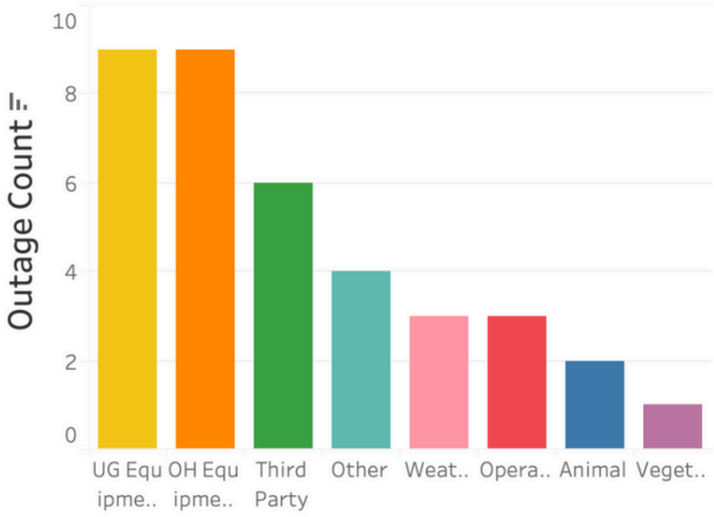


Seasonal Outage Causes - Dashboard

Winter

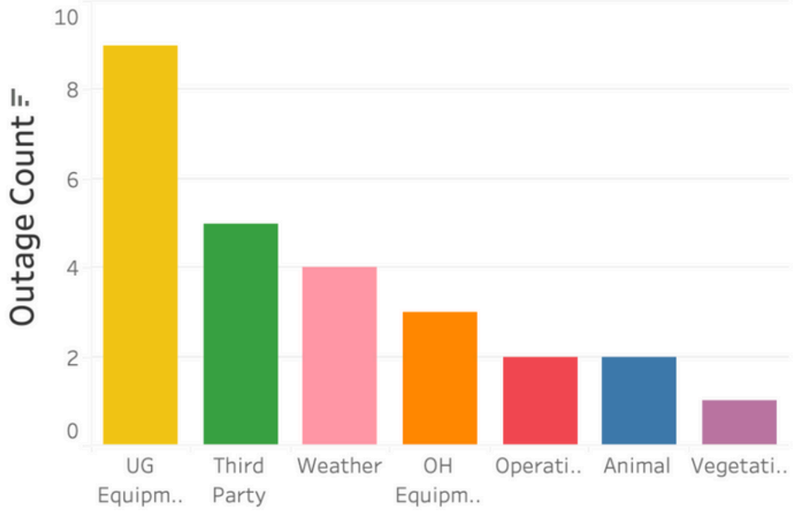


Spring

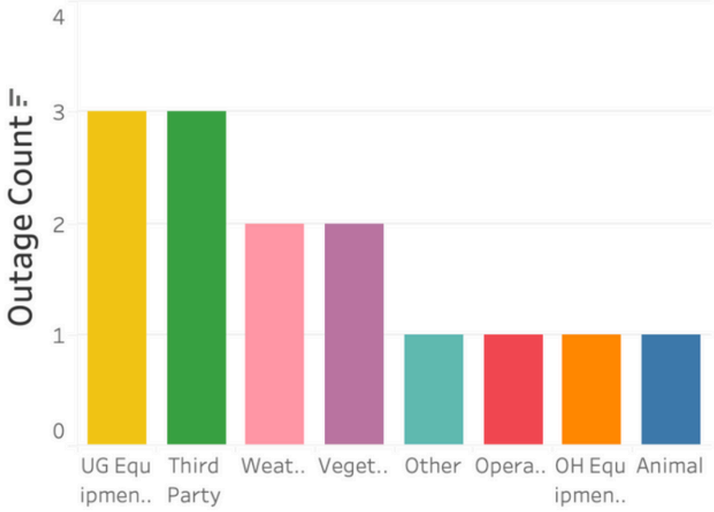


Season	Outage Count	Total CMI
Fall	14	9,914,374
Spring	37	24,202,316
Summer	26	24,039,742
Winter	23	17,813,559

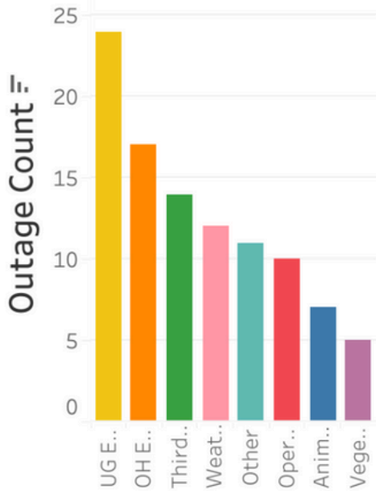
Summer



Fall



2024



Key Insights

- Further identify the leading contributors of outages.
- Magenta: 3 out of 10 outages by OH equipment failure.
- *Roosevelt: 4 of its 9 outages caused by UG equipment failure.*
- *Gorilla: 4 out of 9 outages caused by OH equipment Failure.*
- *Blue Jay, Grand, and Lightning circuits only had 1 outage*

Desert Region

region	circuit_name	outage_count	top_outage_cause	cause_proportion
Desert	Gorilla	9	OH Equipment Failure	0.44
Desert	Oregon	4	Weather	0.50
Desert	Adams	2	Operation	0.50
Desert	Monterey	2	Other	0.50
Desert	Blue Jay	1	Animal	1.00

Coastal Region

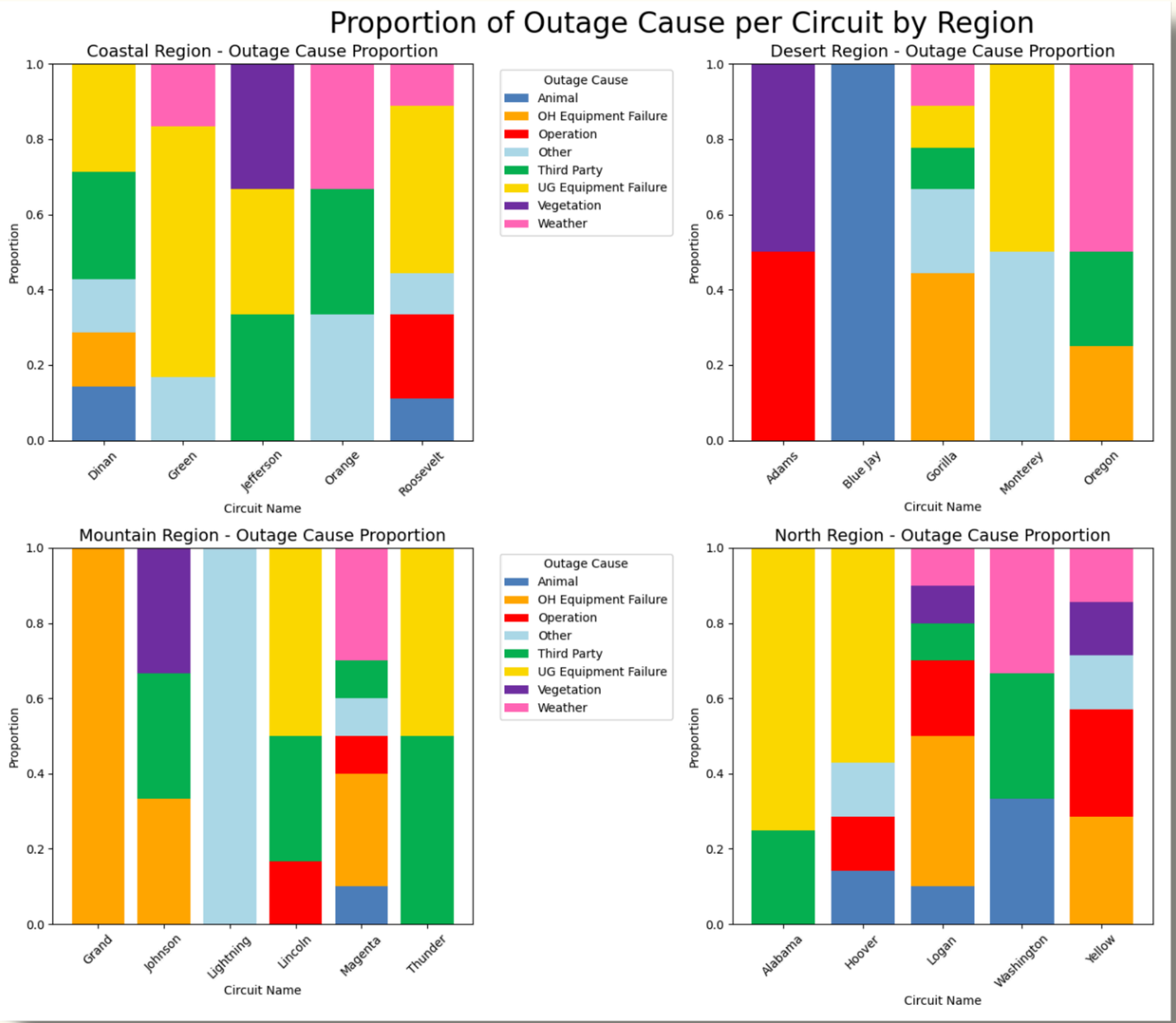
region	circuit_name	outage_count	top_outage_cause	cause_proportion
Coastal	Roosevelt	9	UG Equipment Failure	0.44
Coastal	Dinan	7	Third Party	0.29
Coastal	Green	6	UG Equipment Failure	0.67
Coastal	Jefferson	3	Third Party	0.33
Coastal	Orange	3	Other	0.33

Mountain Region

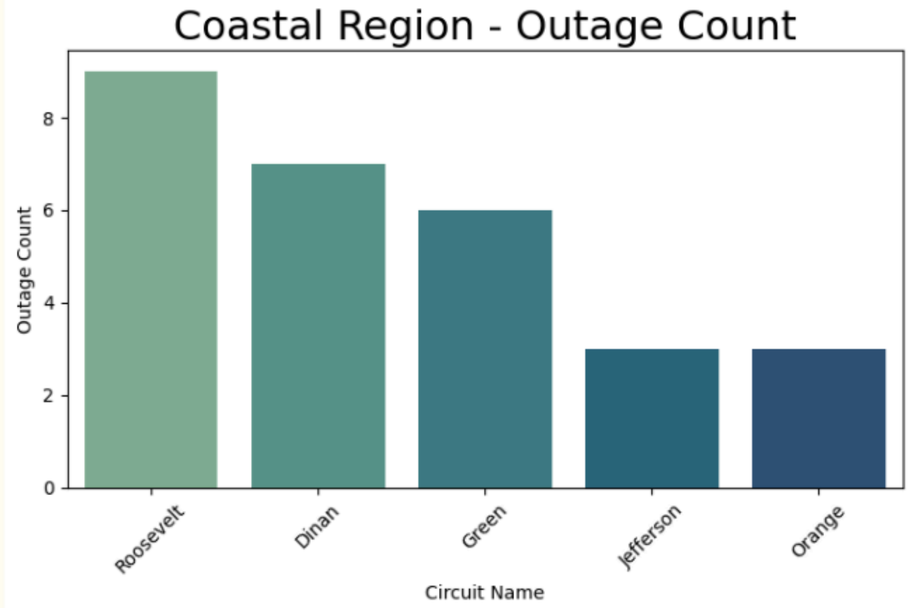
region	circuit_name	outage_count	top_outage_cause	cause_proportion
Mountain	Magenta	10	OH Equipment Failure	0.30
Mountain	Lincoln	6	UG Equipment Failure	0.50
Mountain	Johnson	3	OH Equipment Failure	0.33
Mountain	Thunder	2	Third Party	0.50
Mountain	Grand	1	OH Equipment Failure	1.00
Mountain	Lightning	1	Other	1.00

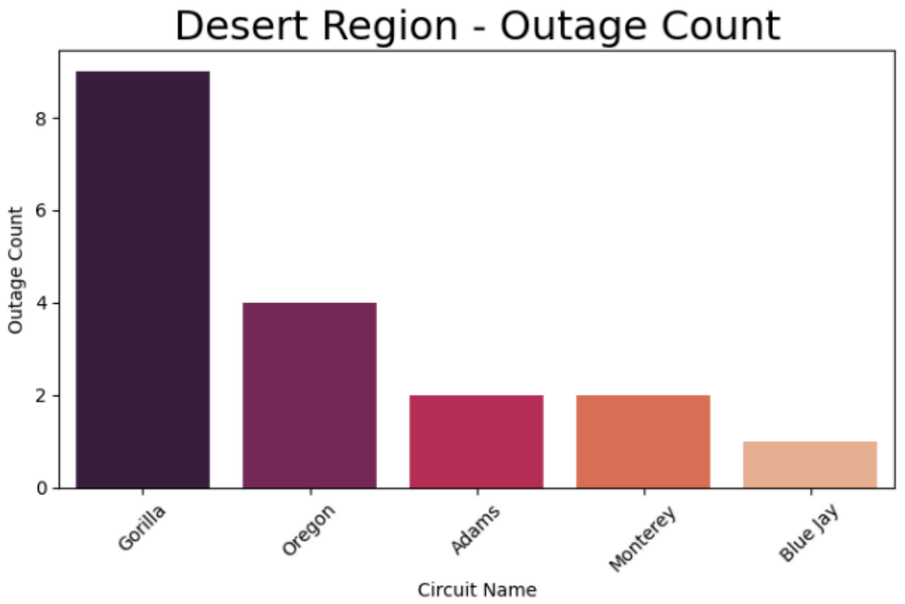
North Region

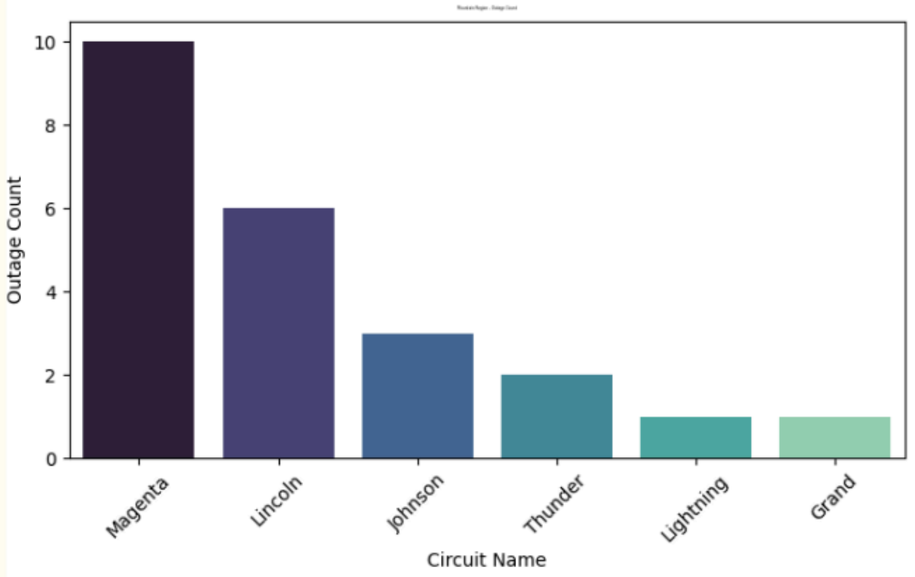
region	circuit_name	outage_count	top_outage_cause	cause_proportion
North	Logan	10	OH Equipment Failure	0.40
North	Hoover	7	UG Equipment Failure	0.57
North	Yellow	7	OH Equipment Failure	0.29
North	Alabama	4	UG Equipment Failure	0.75
North	Washington	3	Animal	0.33

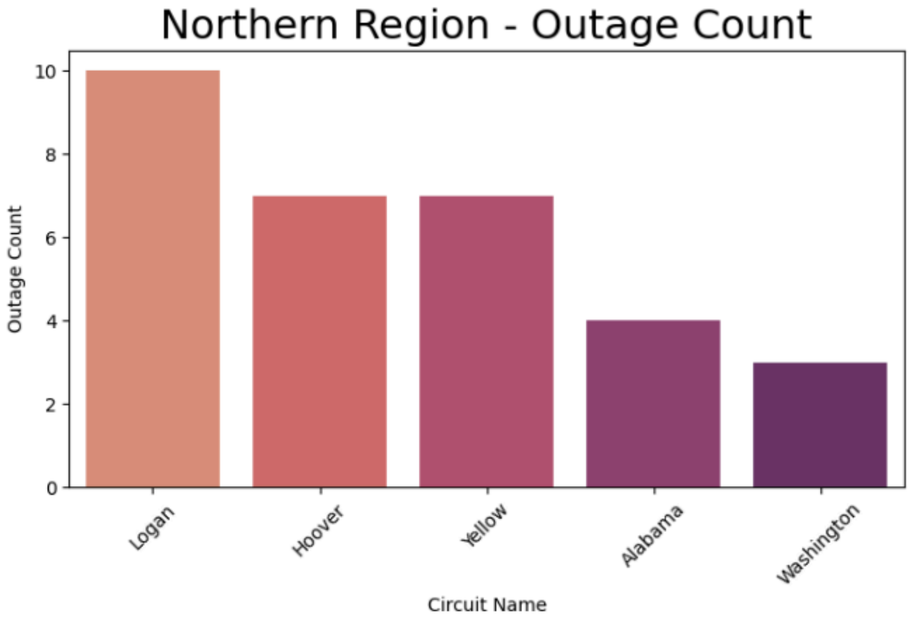


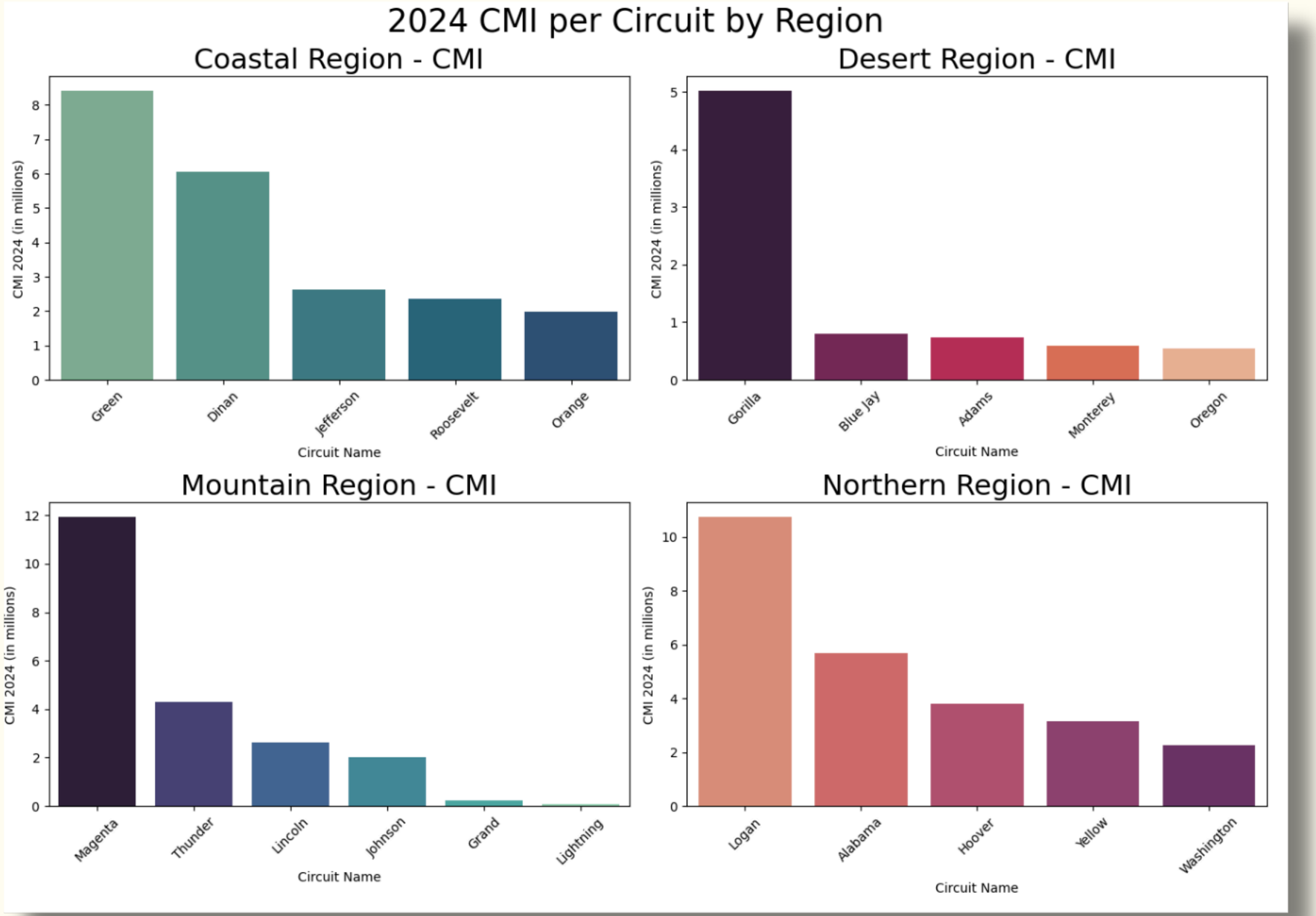
2024 Outage Count per Circuit by Region





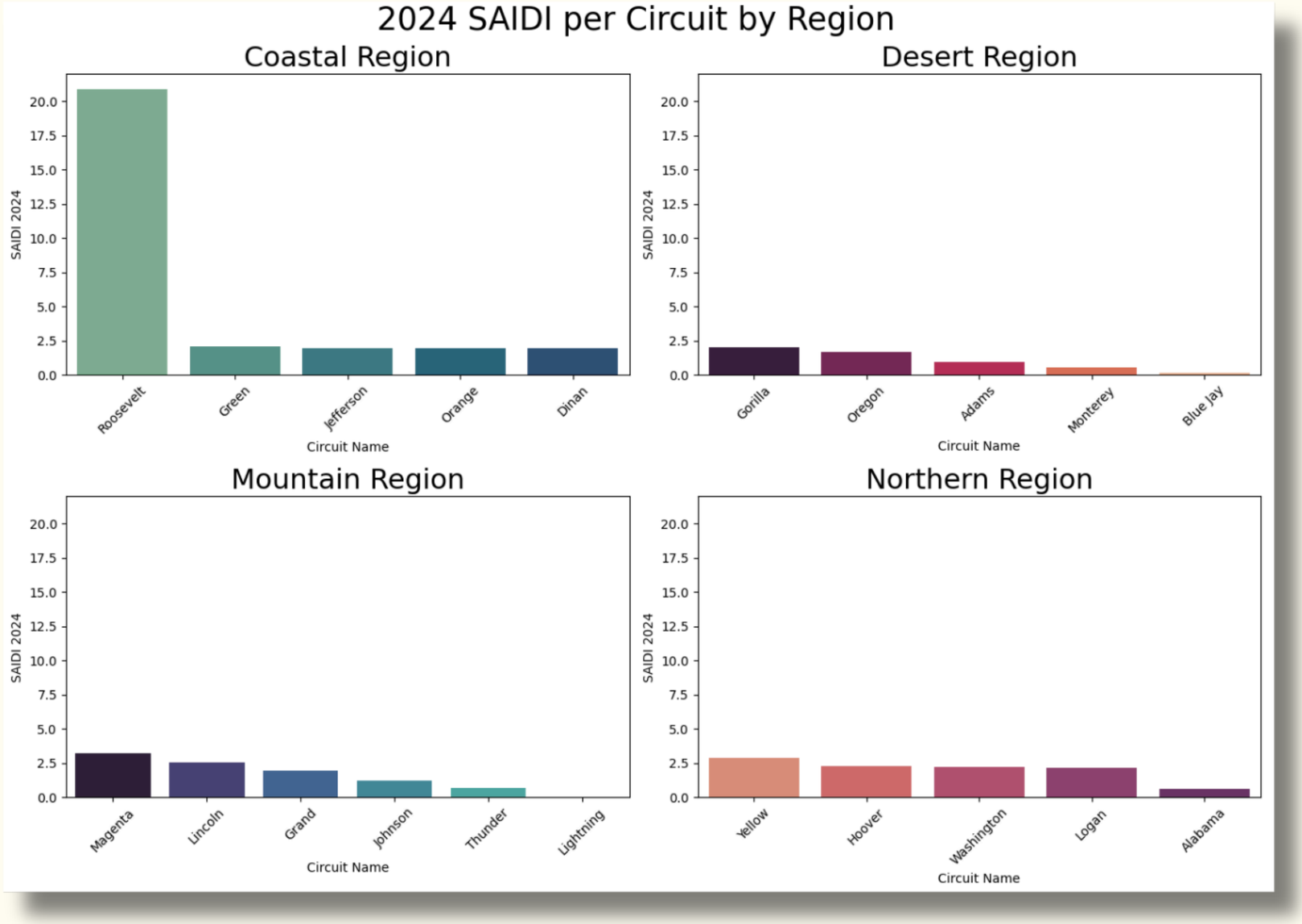


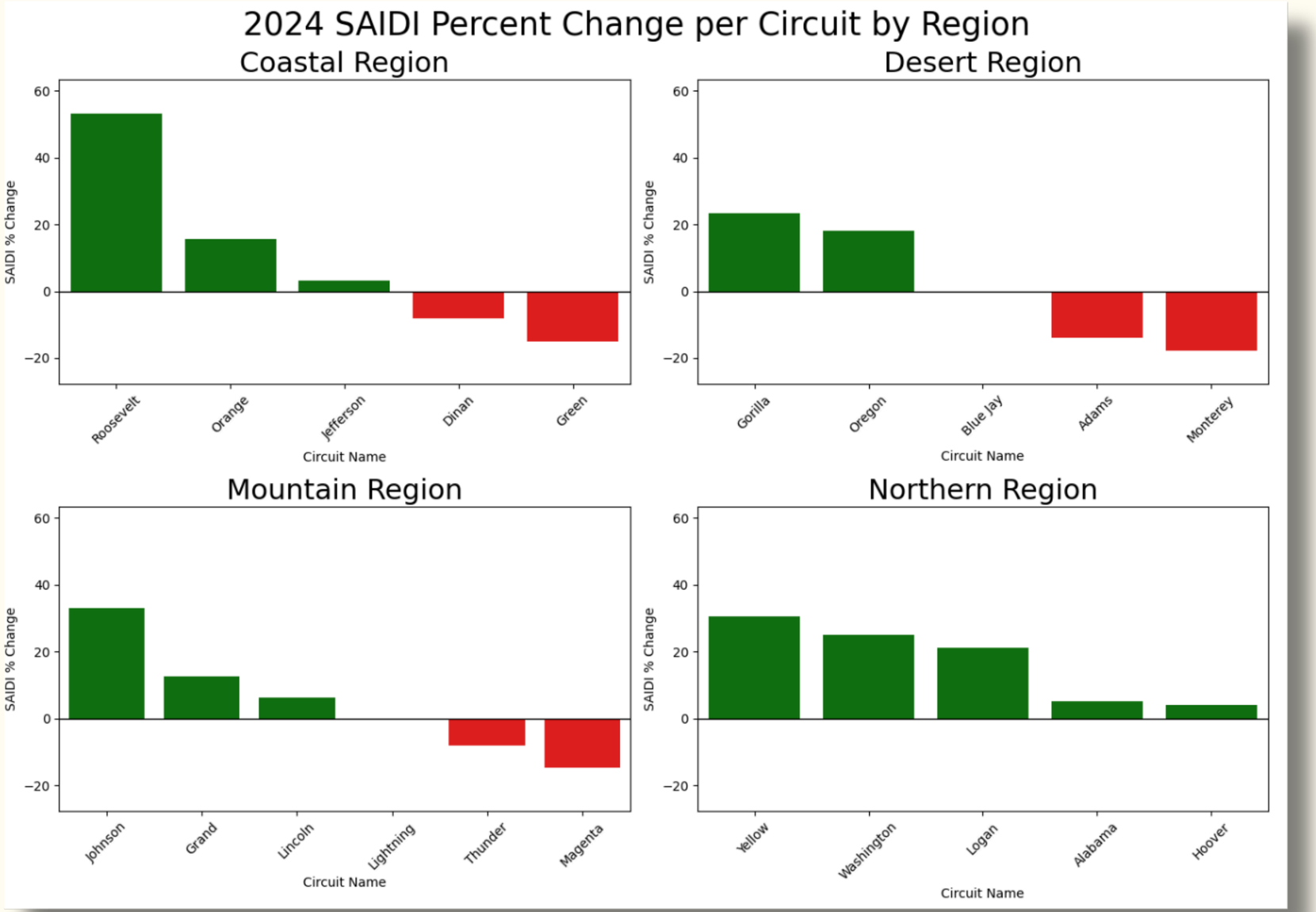



2024 CMI per Circuit by Region
Coastal Region - CMI
CMI 2024 (in millions)
Green
Dinan
Jefferson
Roosevelt
Orange
Circuit Name
Desert Region - CMI
CMI 2024 (in millions)
Gorilla
Blue Jay
Adams
Monterey
Oregon
Circuit Name
Mountain Region - CMI
CMI 2024 (in millions)
Magenta
Thunder
Lincoln
Johnson
Grand
Lightning
Circuit Name
Northern Region - CMI
CMI 2024 (in millions)
Logan
Alabama
Hoover
Yellow
Washington
Circuit Name

SAIDI = total outage duration / customer count

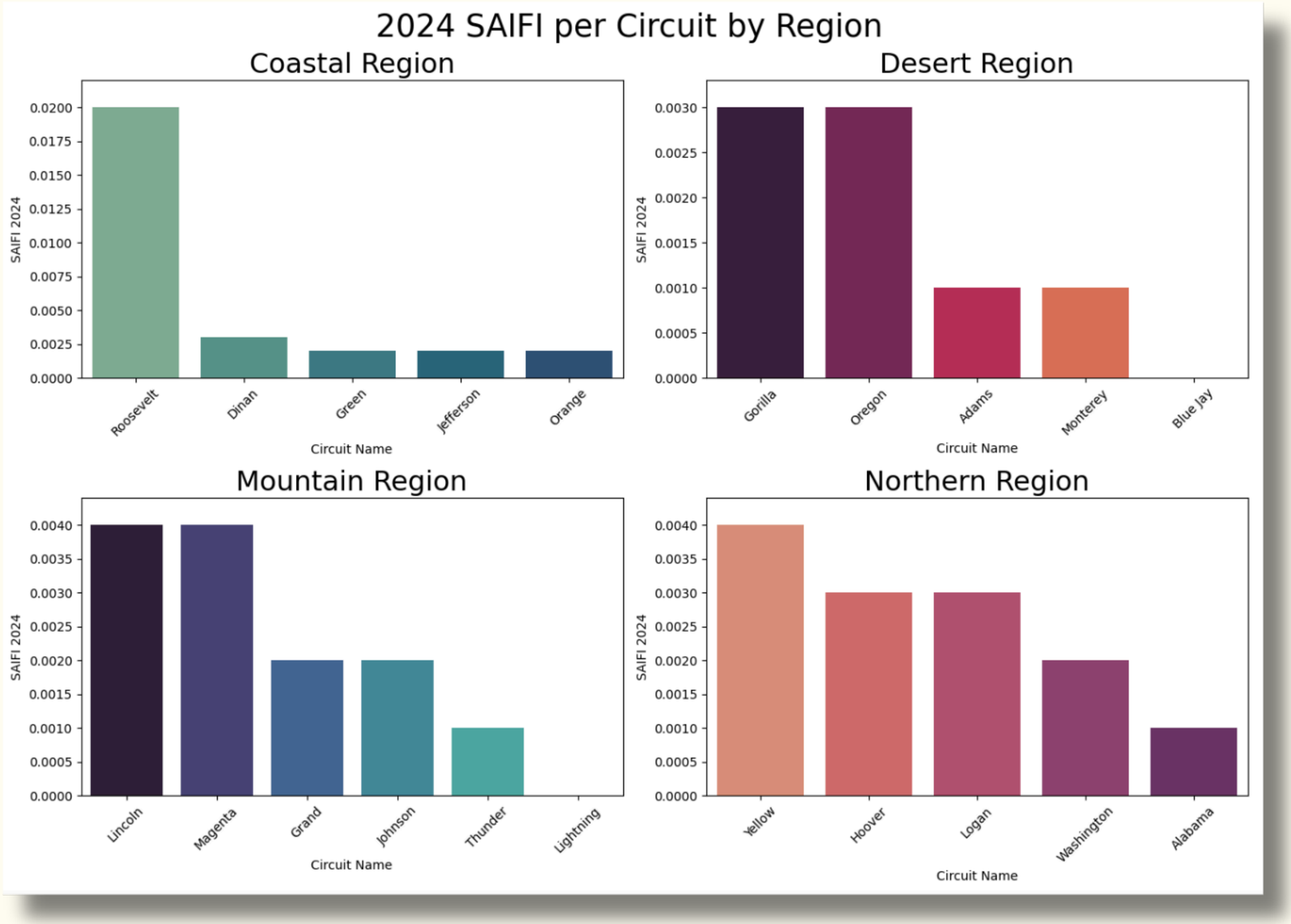
- Roosevelt SAIDI: 20.92
- All other circuits had SAIDI < 3.0

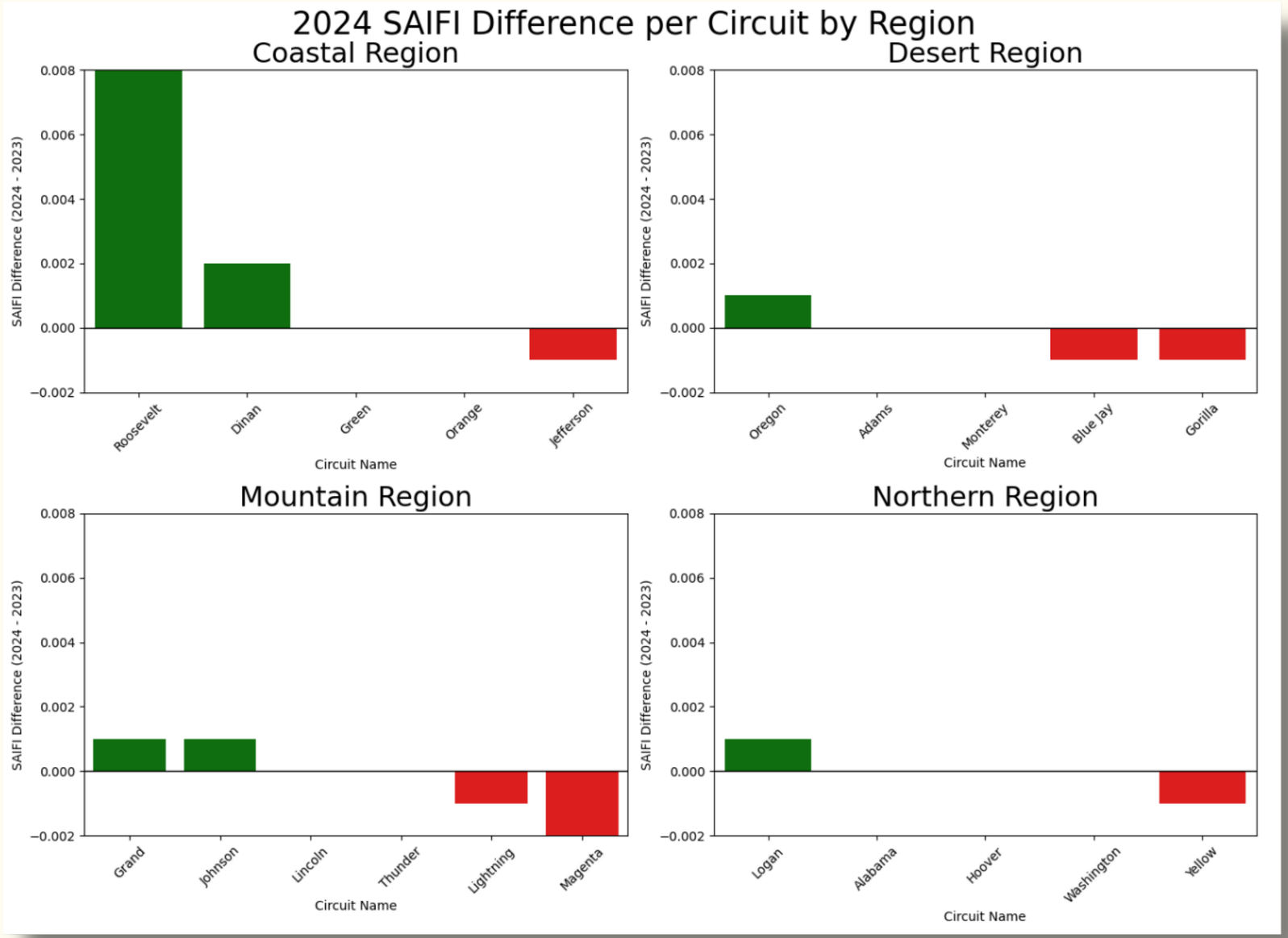


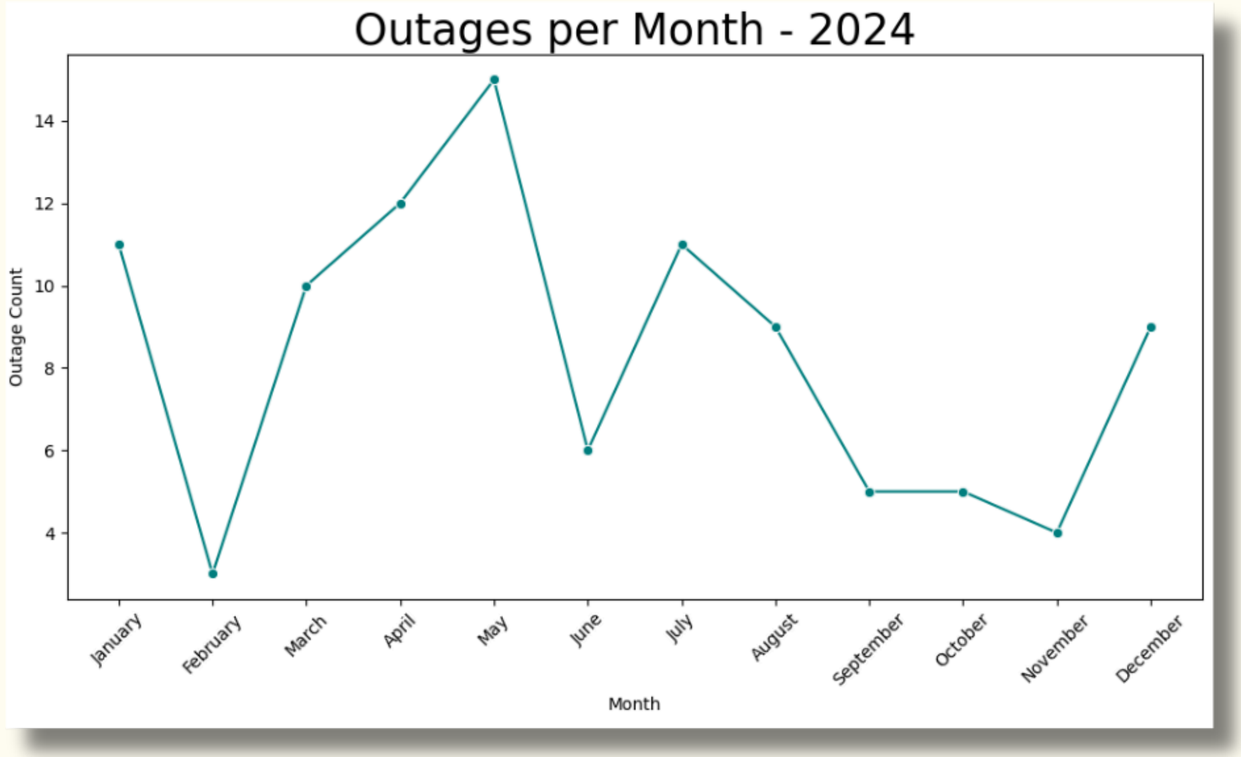

2024 SAIDI Percent Change per Circuit by Region
Coastal Region
Desert Region
Mountain Region
Northern Region
SAIDI % Change
Circuit Name
60
40
20
0
−20
Roosevelt
Orange
Jefferson
Dinan
Green
Gorilla
Oregon
Blue Jay
Adams
Monterey
Johnson
Grand
Lincoln
Lightning
Thunder
Magenta
Yellow
Washington
Logan
Alabama
Hoover

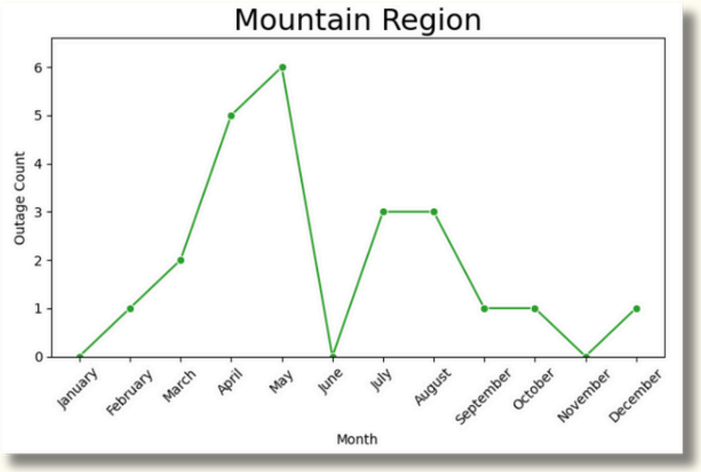
SAIFI= total outages / customer count

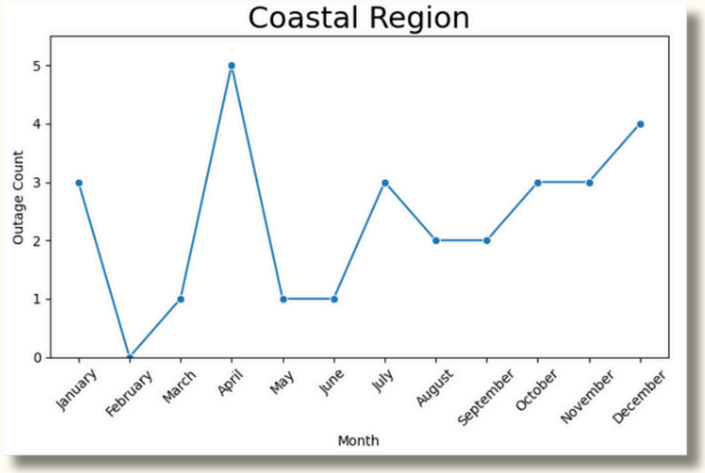
- Roosevelt SAIFI: 0.02
- All other circuits had SAIFI < 0.040

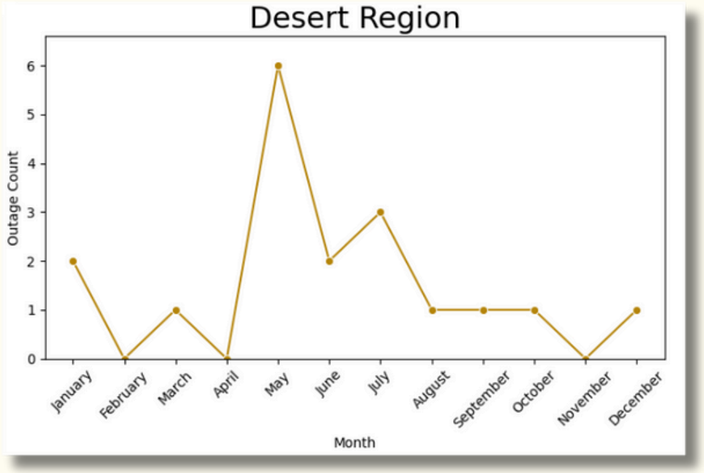


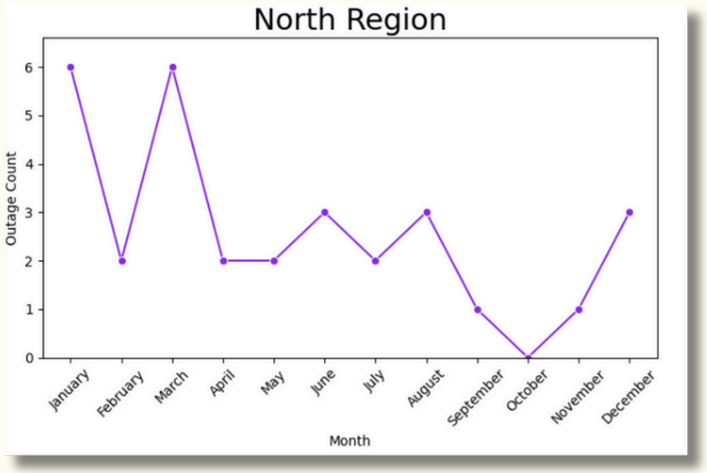

2024 SAIFI Difference per Circuit by Region
Coastal Region
Desert Region
Mountain Region
Northern Region
SAIFI Difference (2024 - 2023)
0.008
0.006
0.004
0.002
0.000
−0.002
Roosevelt
Dinan
Green
Orange
Jefferson
Oregon
Adams
Monterey
Blue Jay
Gorilla
Grand
Johnson
Lincoln
Thunder
Lightning
Magenta
Logan
Alabama
Hoover
Washington
Yellow
Circuit Name

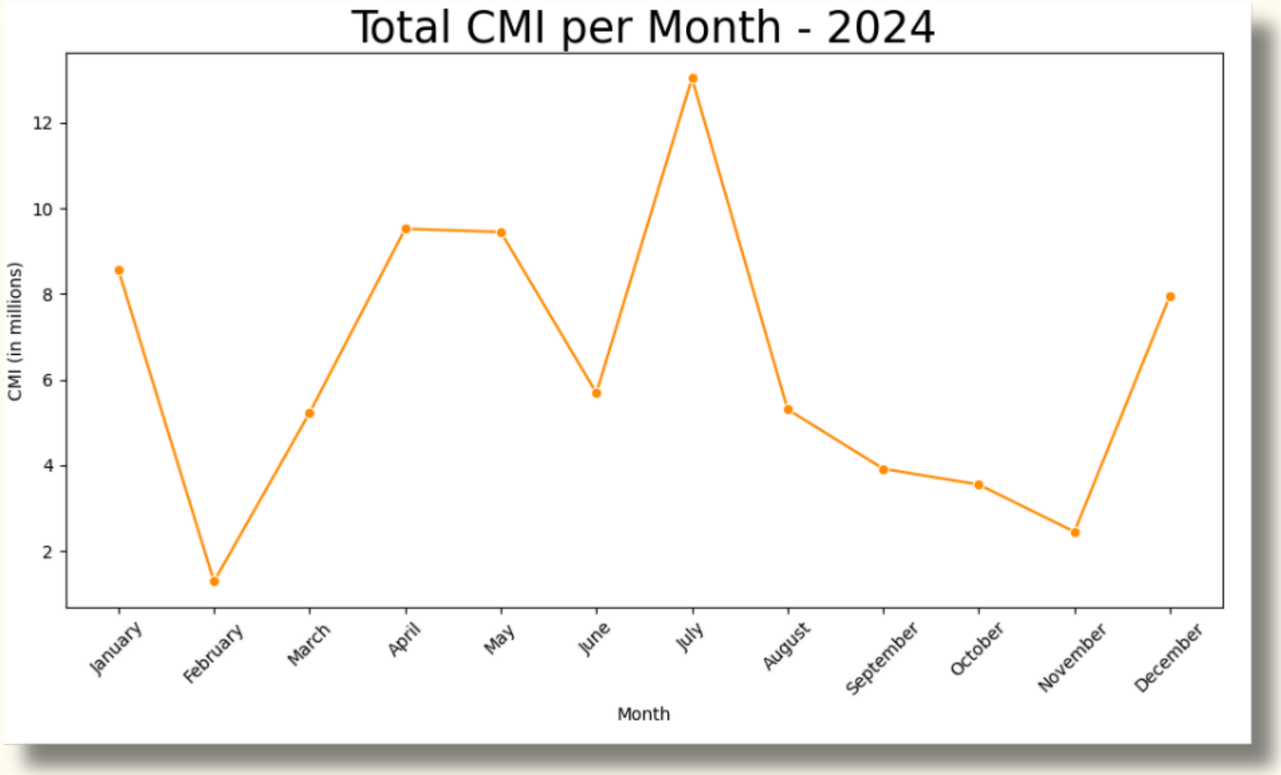

Outages per Month - 2024
Outage Count
Month
January
February
March
April
May
June
July
August
September
October
November
December

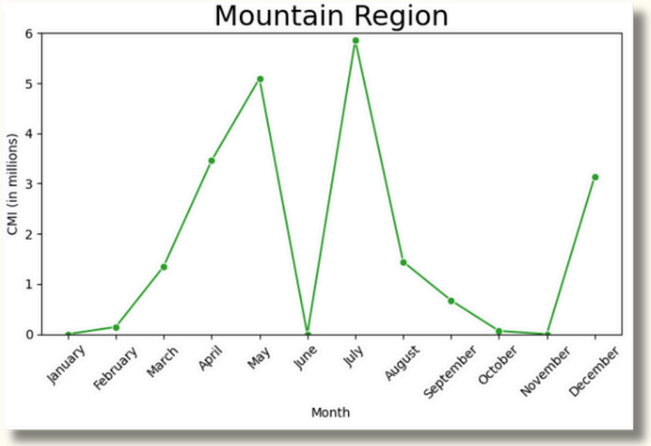

Mountain Region
Outage Count
Month

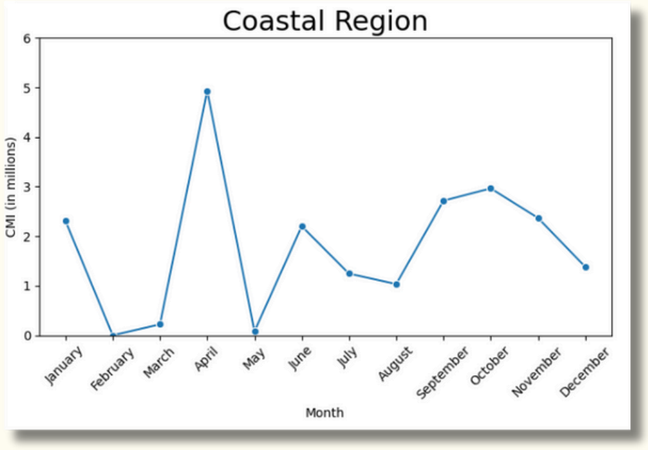

Coastal Region
Outage Count
Month

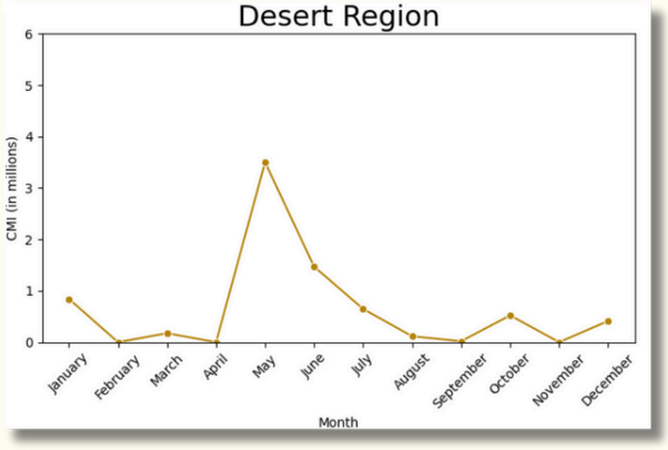

Desert Region
Outage Count
Month

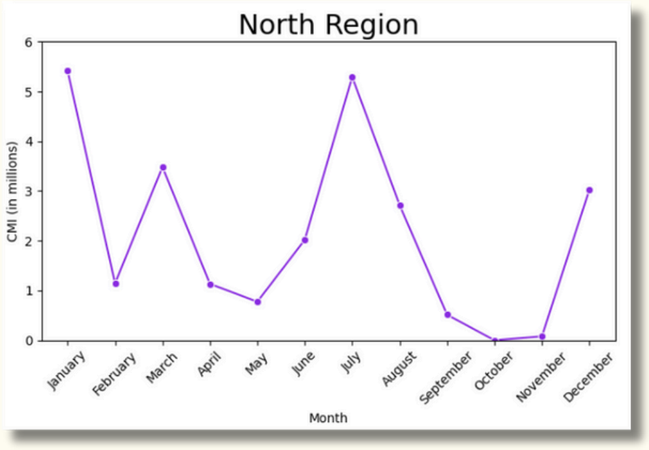

North Region
Outage Count
Month


Total CMI per Month - 2024
CMI (in millions)
Month
January
February
March
April
May
June
July
August
September
October
November
December


Mountain Region
CMI (in millions)
Month


Coastal Region
CMI (in millions)
Month

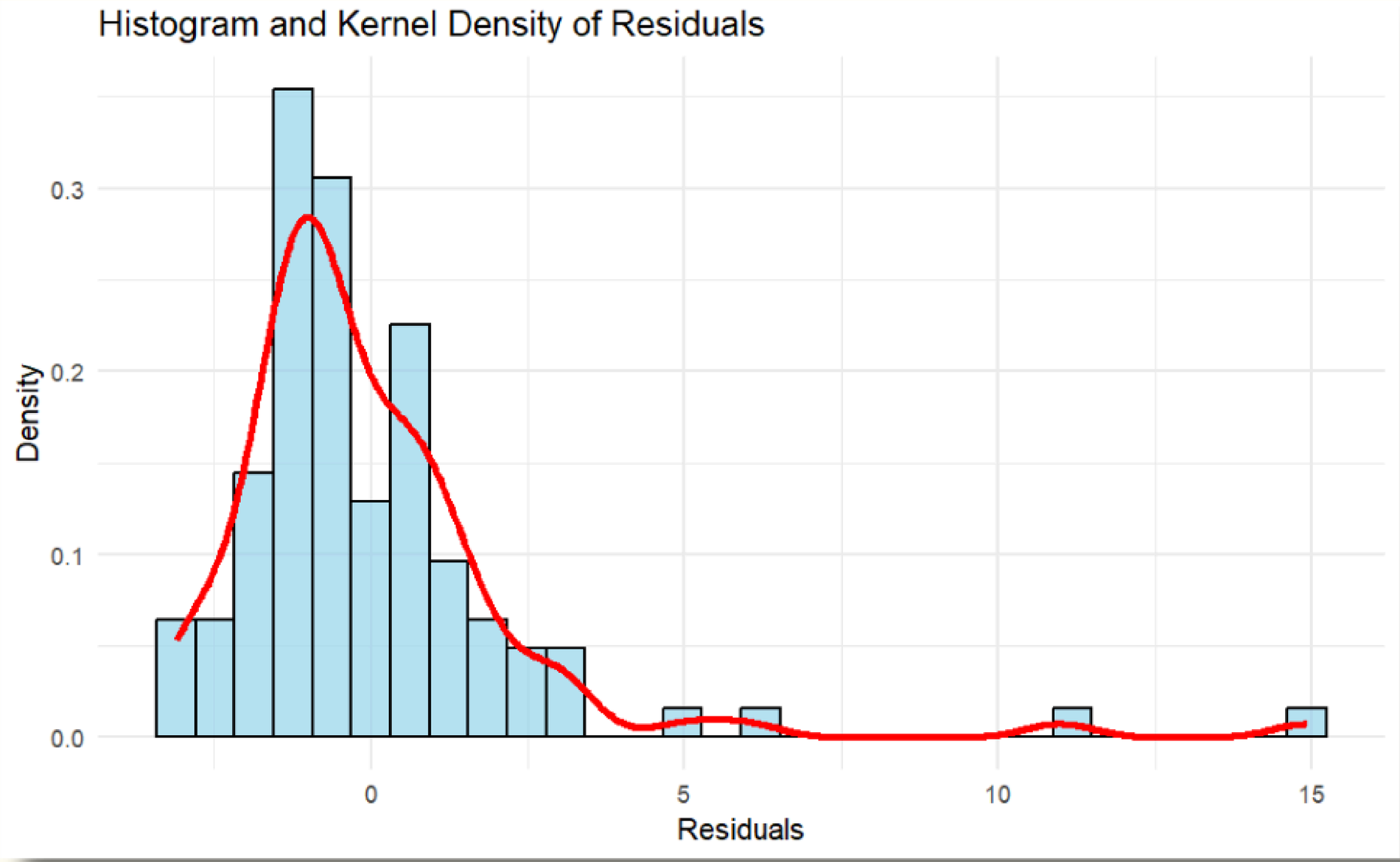

Desert Region
CMI (in millions)
Month


North Region
CMI (in millions)
Month

MULTIPLE LINEAR REGRESSION (MLR)

We can use **Multiple Linear Regression (MLR)** on *average wait time per customer* to forecast different cases scenarios that fit within our interpolation.

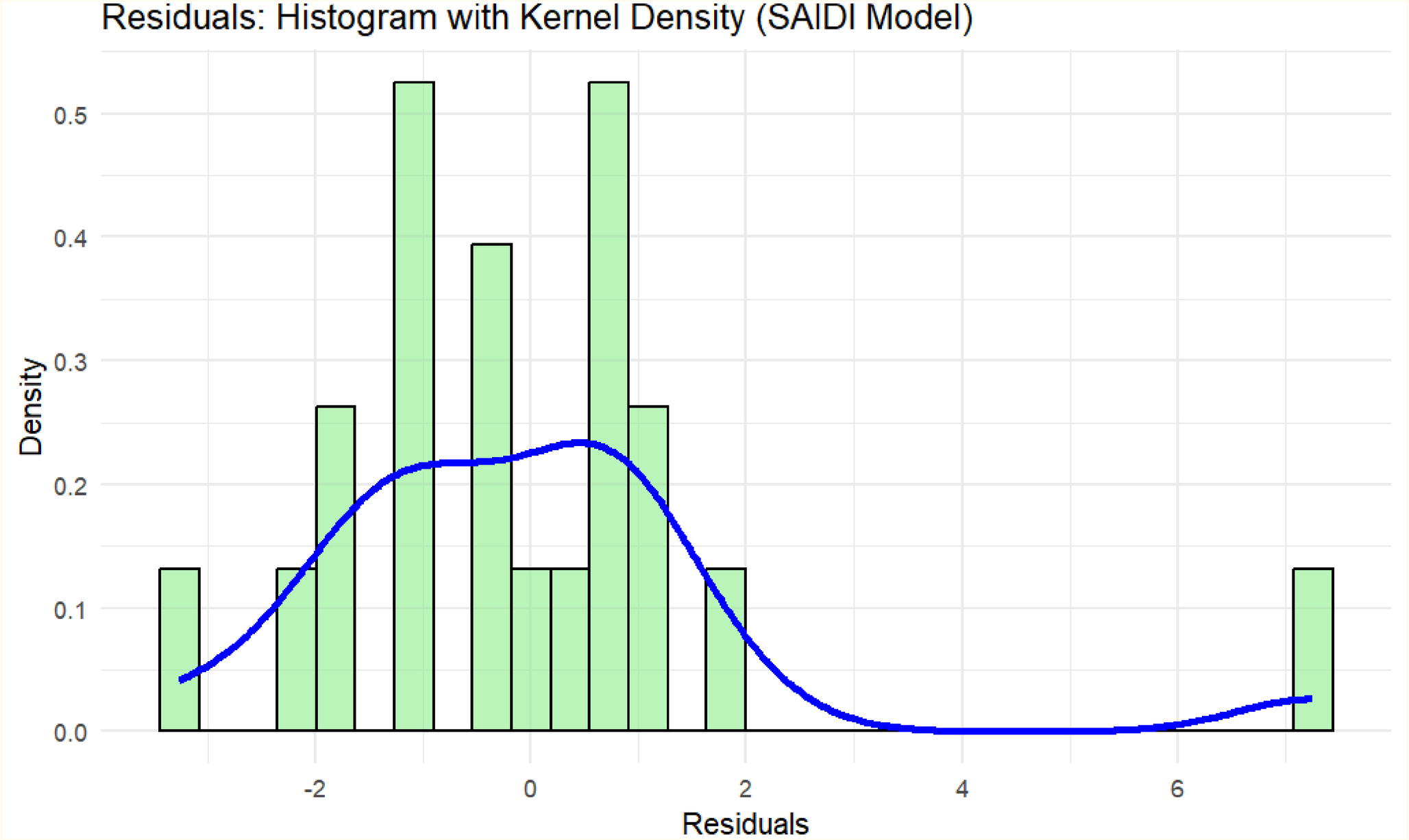
We get a **Multiple R-square value of 0.4083**, so 40.83% of the variance of *average wait time per customers* can be explained by the independent variables.


Histogram and Kernel Density of Residuals
Density
0.0
0.1
0.2
0.3
0
5
10
15
Residuals

MLR on *System Average Interruption Duration Index (SAIDI)*

Using the **secondary dataset,** we model **SAIDI** to potential forecast hypothetical situations of how different outage scenarios might affect system dependibility.

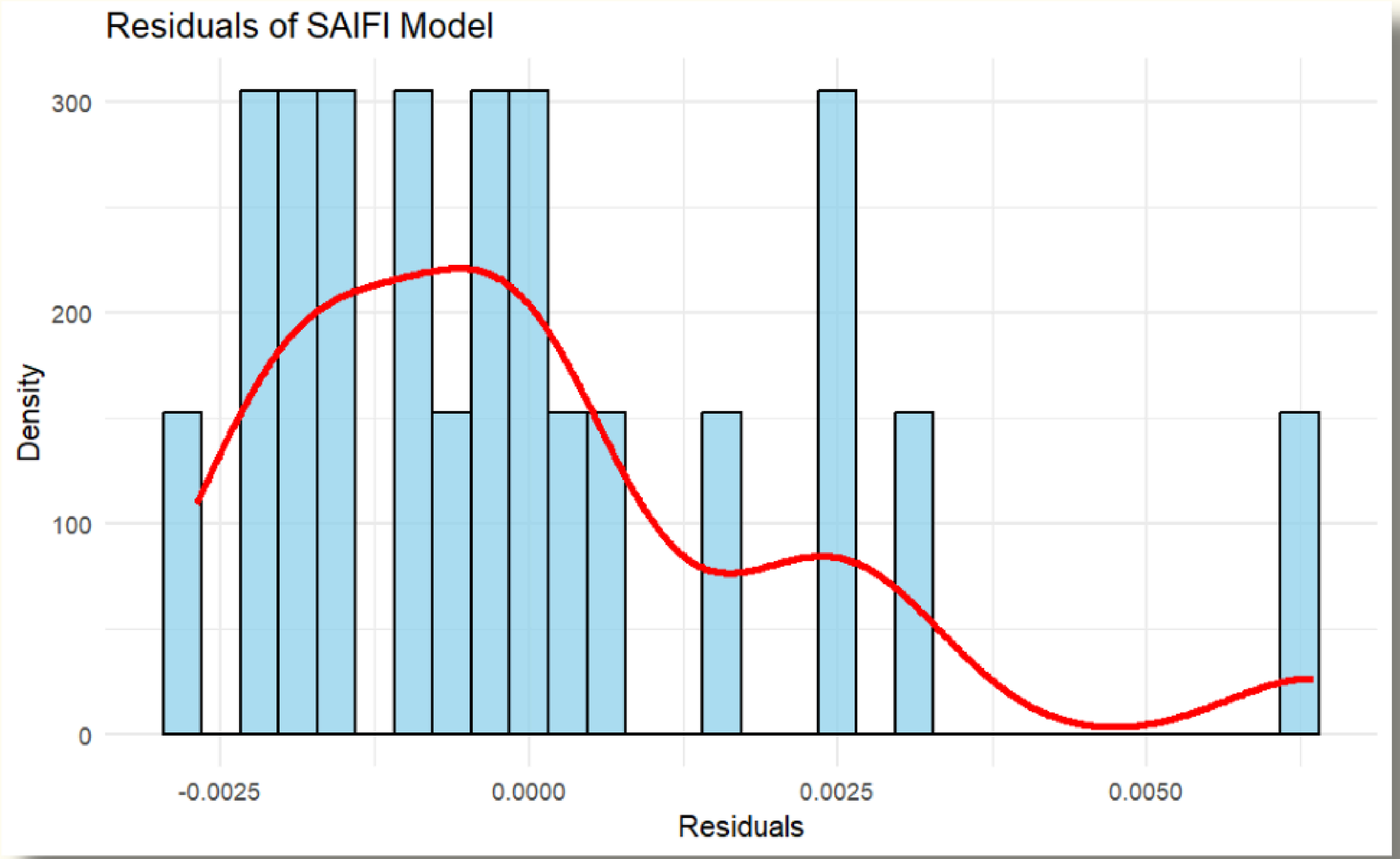
Our model achieved a **Multiple R-Squared value of 0.428**, which means 42.8% of variation in SAIDI can get explained by the independent variables used in the model.


Residuals: Histogram with Kernel Density (SAIDI Model)
Density
0.0
0.1
0.2
0.3
0.4
0.5
-2
0
2
4
6
Residuals

MLR on *Systems Average Interruption Frequency Index (SAIFI)*

Using the **secondary dataset**, we modeled **SAIFI** to forecast hypothetical scenarios and predict how different outage conditions might impact system performance.

Our model produced a **Multiple R-Squared value of 0.3007**, meaning 30.07% of the variation in SAIFI can be explained by the independent variables used.


Residuals of SAIFI Model
300
200
100
0
Density
-0.0025
0.0000
0.0025
0.0050
Residuals

Logistic Regression

Logistic Regression Findings (95% Confidence Level)

- **Holiday**: Outage probability between **15.42% and 37.15%**
- **Post-Holiday**: Outage probability between **24.16% and 68.00%**, with **17.46% higher odds** compared to holidays.
- **Regular Day**: Outage probability between **33.02% and 74.82%**, with **21.05% higher odds** compared to holidays.

Key Insight

Outages are more likely to occur on *regular days* than on holidays, suggesting an opportunity for IBEC to **optimize crew scheduling** around calendar patterns.

	YES	NO	Total
Holiday	**3**	**11**	**14**
Post-holiday	**4**	**10**	**14**
Regular day	**84**	**254**	**338**
Total	**91**	**275**	**366**

How does Multiple Linear Regression help IBEC?

We used predictive modeling to simulate different **“what-if”** scenarios, helping IBEC estimate how improvements in **wait time**, **SAIFI**, or **SAIDI** could justify infrastructure investments.

Our analysis also showed that outages are **more likely** on *regular days* compared to holidays, meaning IBEC can better plan crew schedules to **respond faster**.

With more long-term outage data, we could further enhance these predictions and **calculate daily outage probabilities**, allowing IBEC to **optimize dispatch staffing** even more effectively.

RECOMMENDATIONS & STRATEGY

PRIORITY ACTIONS

HIGHER PRIORITY

Action	Strategic Outcome
Upgrade Critical Circuits	Fastest CMI reduction and improved customer reliability
Expand Predictive Maintenance Across At-Risk Areas	Proactively prevents outages and reduces reactive repair costs
Modernize Grid with Smart Technologies	Accelerates outage detection and strengthens system resilience
Strengthen Vegetation and Environmental Risk Controls	Minimizes wildfire and weather-related disruptions
Enhance Outage Data and Dashboards	Empowers real-time decision-making and regulatory compliance
Prepare for Strategic Undergrounding and Data Integration	Future-proofs the grid against environmental risks and system failures

STRONGER ROI

SHORT-TERM STRATEGY (Q2–Q4 2025)- IMMEDIATE ACTION

Inspect

- Targeted inspections of overhead equipment on the ***Magenta*** and ***Logan*** circuits (early signs of wear, damage, recurring faults)
- ***Roosevelt*** circuit diagnostic initiative to *investigate* high SAIDI/SAIFI values

Monitor

- **Thermal imaging** inspections (May-June) to *detect* underground component overheating
- **Flood zone mapping** using FEMA data to *identify* UG areas at risk

Develop

- Pilot a **reliability dashboard** for planners/engineers to monitor CMI and SAIDI
- Establish internal standards for how **outage events** are *recorded*, *categorized*, and *validated*

Implement

- **Install** fault detection sensors on ***Green*** and ***Thunder*** circuits
- Deploy *passive cooling systems* or *imported conduit ventilation*
- *LiDar*-based vegetation management in *Mountain* and *North* regions **before spring**

LONG-TERM STRATEGY (2026–2028) – SYSTEM IMPROVEMENTS

Maintenance Expansion

- **Expand** *predictive maintenance efforts* to cover additional circuits showing early signs of *declining reliability*, as measured by **rising** SAIDI/SAIFI trends.

Grid Modernization

- **Invest** in *remote monitoring* and *auto-fault* isolation
- Use thermal imaging to **detect** *overheated UG* components
- Incorporate FEMA flood data to **map** UG areas *at risk*

Conversion

- Bring select overhead circuits underground, **prioritizing** areas with *repeated* outages and risk exposure from weather or vegetation.

Integration

- **Integrate** external datasets such as wildfire risk maps, climate forecasts, and FEMA flood zones into IBEC's outage analytics.

LIMITATIONS

- Limited historical data, restricted long-term trend forecasting.
- Lack of geographic precision, not detailed enough to perform spatial risk analysis.
- Synthetic and incomplete external data, external research could not be fully integrated for comparison analysis.
- Model assumptions and accuracy, moderate at-best results indicate room for further refinement.

Thank you!