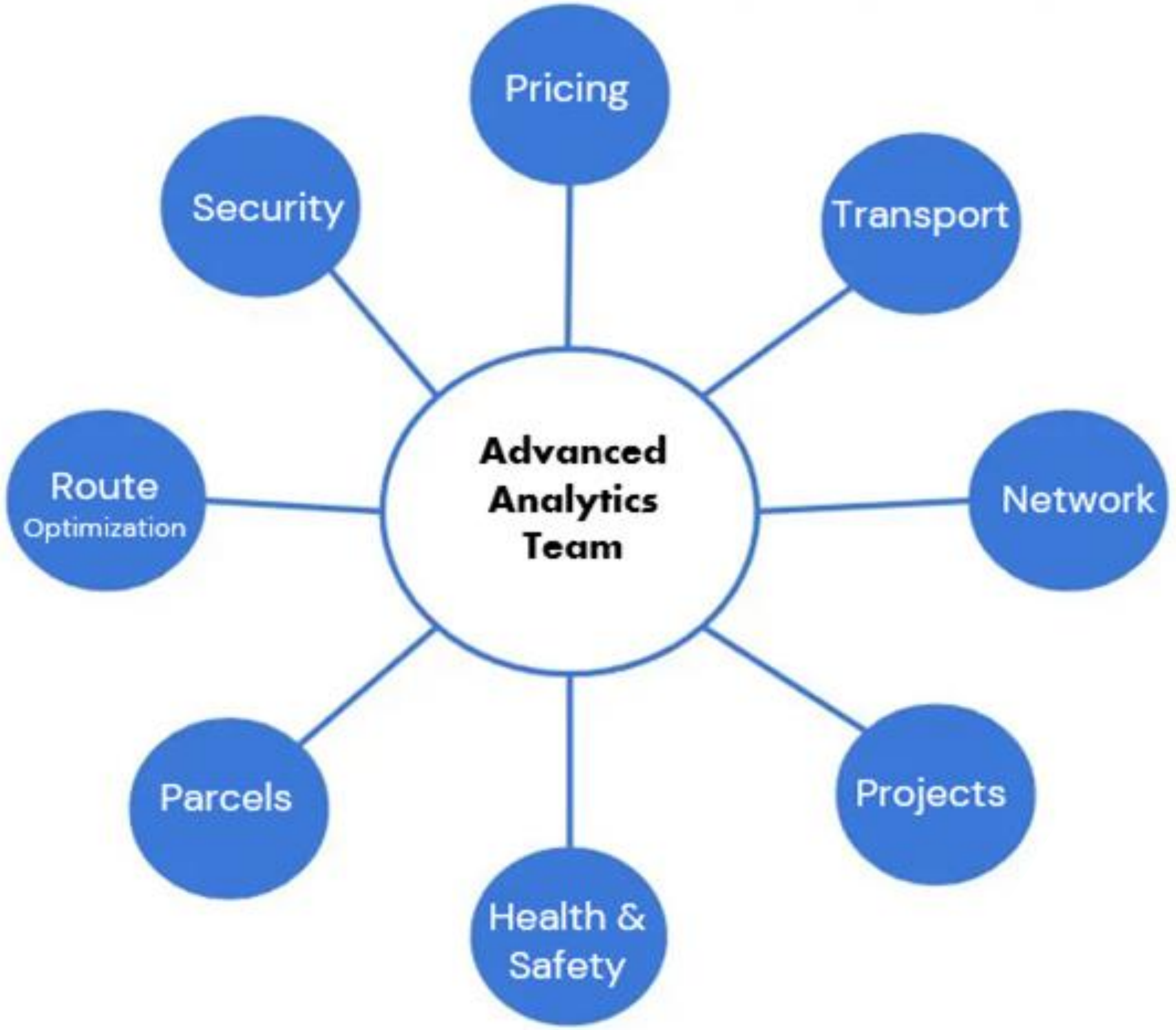
The Evolution of Analytics at Canada Post:
Going Beyond Prediction

Carol Wilson
Director of Advanced Analytics

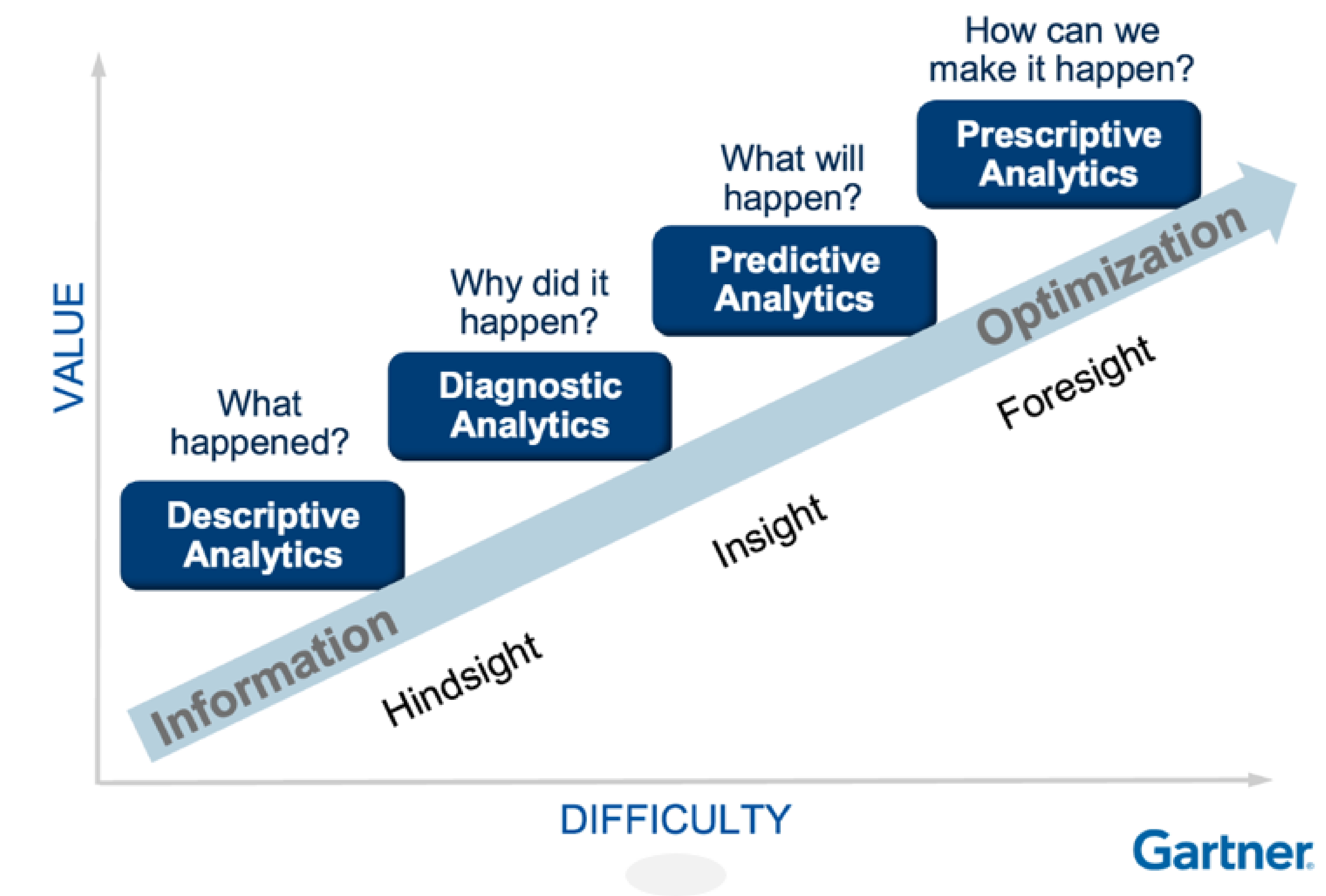
Hub and Spoke Model for Analytics

- Support spokes “Meet them where they are at”
- Educate spokes
- Encourage spokes to conduct own analysis
- Analytics CoE focusses on more complex questions
- Promote our services — Dog and Pony Shows



The Evolution of Analytics at Canada Post: Going Beyond Prediction

Gartner Analytic Ascendancy Model



Descriptive Analytics

Data Literacy

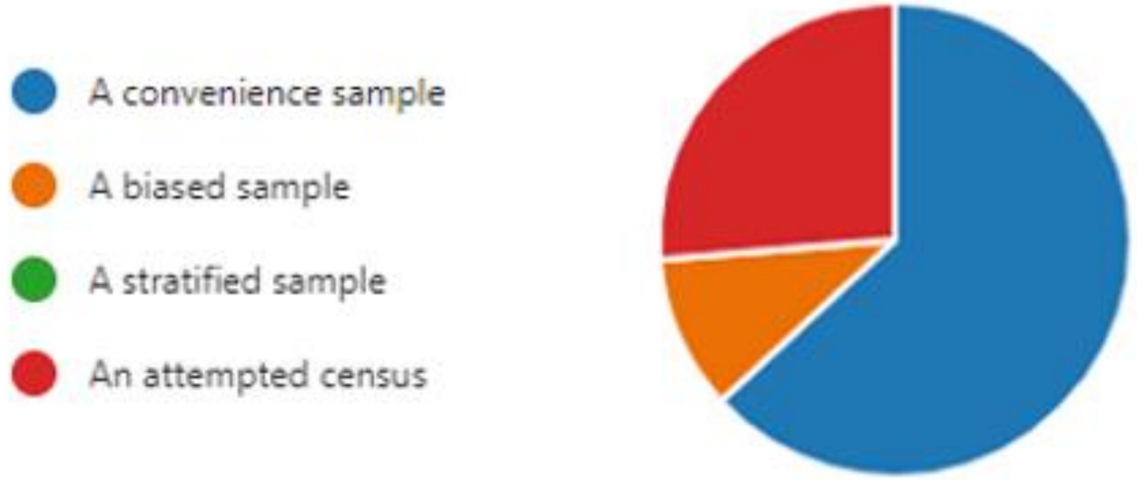
- Need to understand:
    - The type of data you have
        - Census
        - Attempted census
        - Random sample
        - Convenience sample

    - When to use statistical testing

Gap in Sampling Knowledge

Data Scientist recruiting quiz question that nearly everyone gets wrong

8. A company sent out an online satisfaction survey to all of its customers. Only 1200 out of all 7000 customers completed the survey within the time frame. In this case, the data collected would be considered:

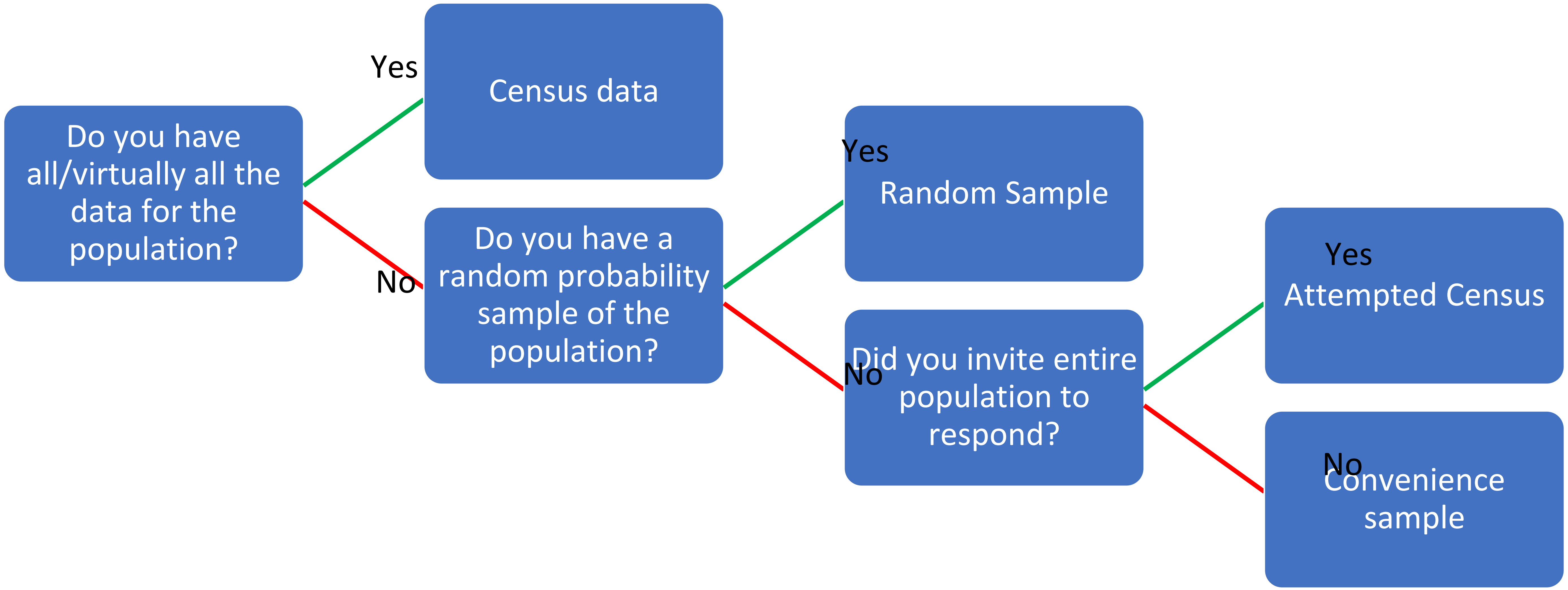


9. In the above example, only about 17% of customers completed the survey. How would this low response rate impact the margin of error for the study?

Determining Differences

- There are always <u>two</u> questions to ask when deciding whether a difference between groups is worth reporting
    1) Is the difference statistically reliable?
    2) Is the difference big enough to be important?
- Test statistics to determine if differences between groups are reliable are <u>only</u> appropriate when you have a **random probability sample**
- With census data we know the data is reliable so the only question is: *"Is the difference big enough to care about?"*

What type of data do I have?



Diagnostic Analytics

Why did it happen?

Diagnostic Analytics with Classification Models

- We use many types of classification models to answer business questions
  - Random Forest, Adaboost, Gradient Boost
  - Detecting Mail Redirection Fraud and other types of fraud
  - What are the biggest risk factors for Letter Carrier injuries

- Both fraud and injury models have an **imbalanced data problem**

- Random Forest
  - Causes of Letter Carrier injuries

Variables Included in Analysis – Potential Drivers of Injury

Mail Volume Features

- Parcel
- Lettermail
- Neighbourhood mail

Employee Features

- Age
- Tenure
- Permanent/Temporary
- Familiarity with Route
- Regular vs overtime hours

Letter Carrier Injuries

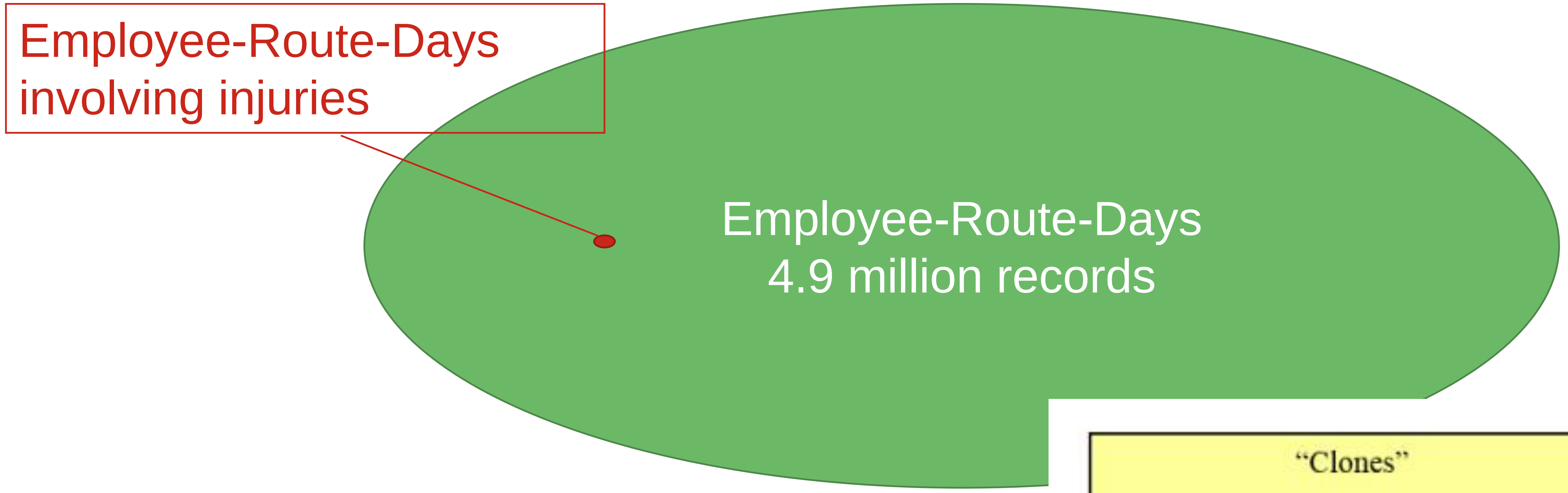
Route Characteristic Features

- Distance Walked
- Business point of calls
- Centralized vs. Door-to-Door
- Number of stairs
- Motorized vs. Foot route

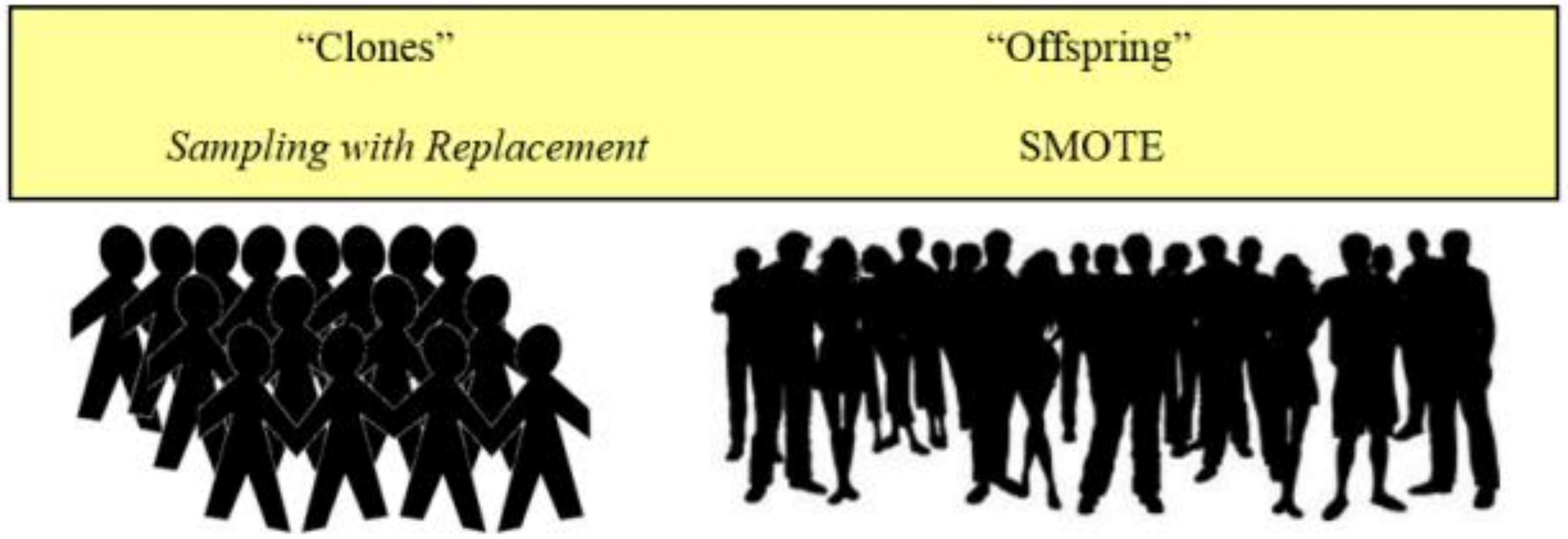
Weather Features

- Maximum temperature for day
- Total of rain for the previous day
- Total of snow for the previous day
- Total rain and snow for delivery day.

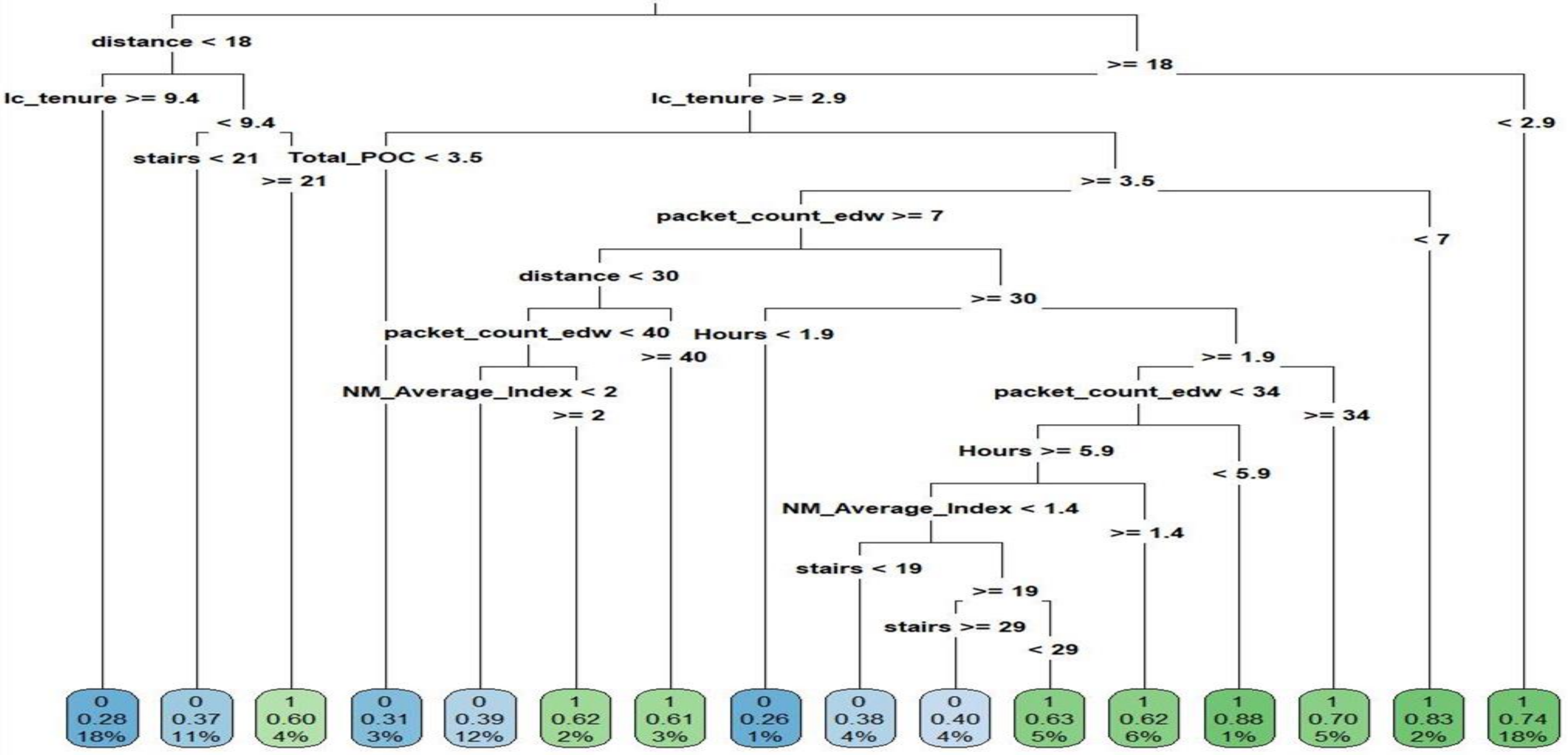
Random Forest Model – Design and Samples



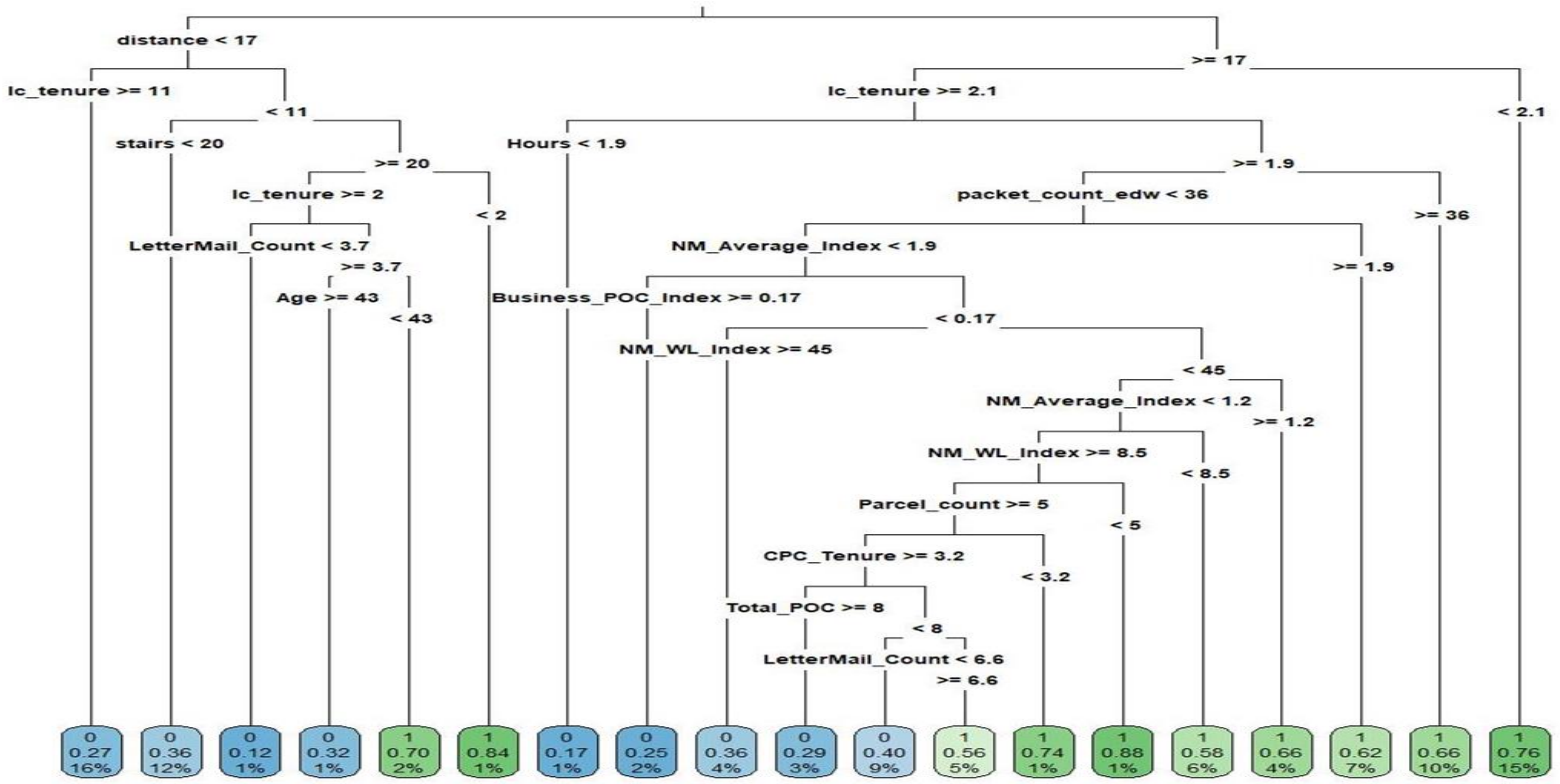
Synthetic Minority Over-Sampling Technique (SMOTE): The SMOTE technique is preferred over *sampling with replacement* because it uses the information available for existing group members to assign reasonable new values to the new sample. Rather than merely replicating existing records in the under-represented population, it is creating new unique members that fit within the logical parameters of the group.



Classic Decision Tree – Random Sample 1

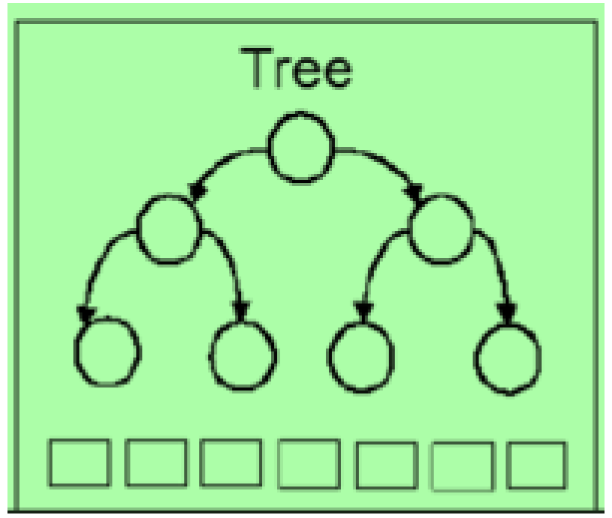


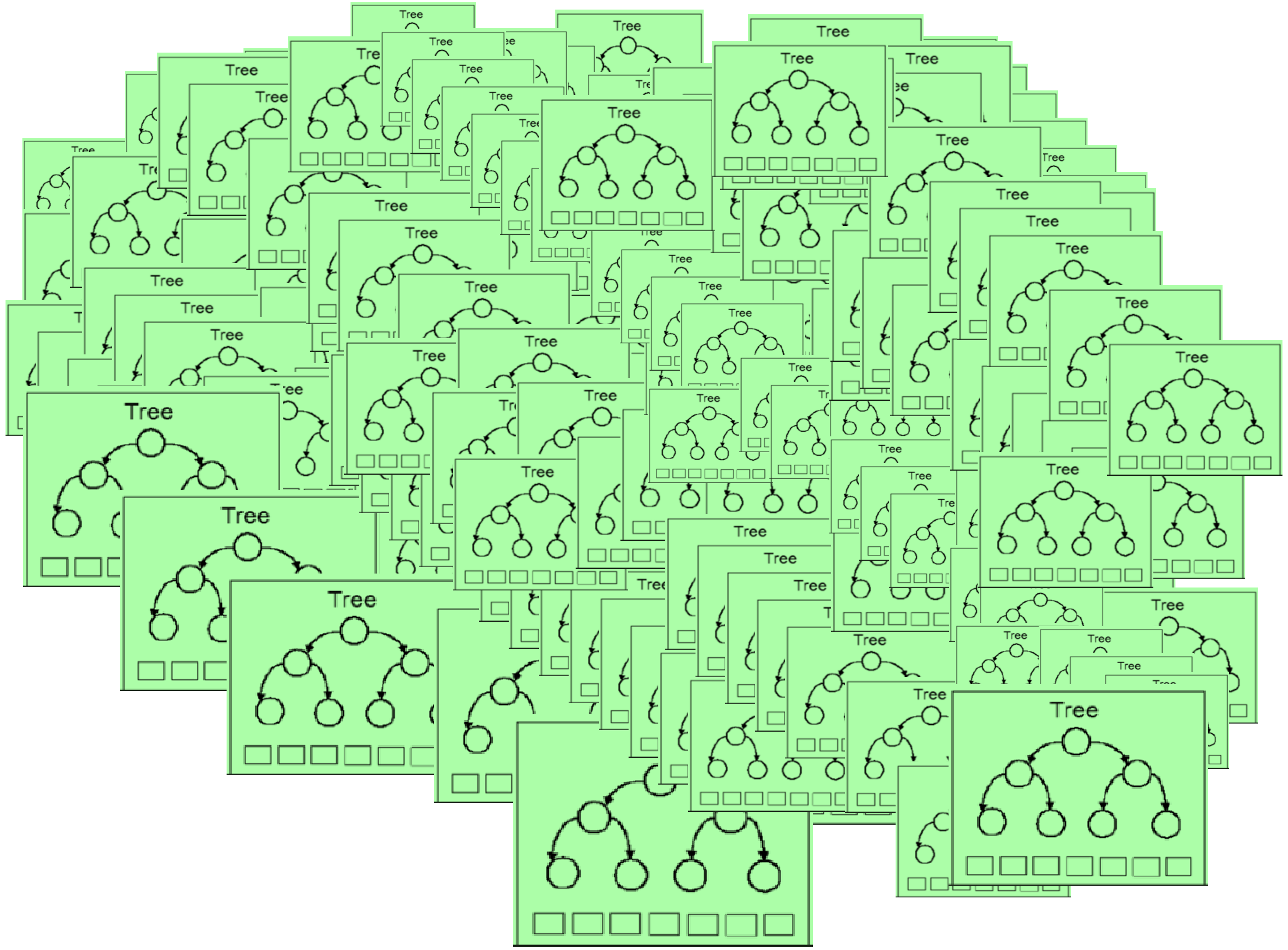
Classic Decision Tree – Random Sample 2



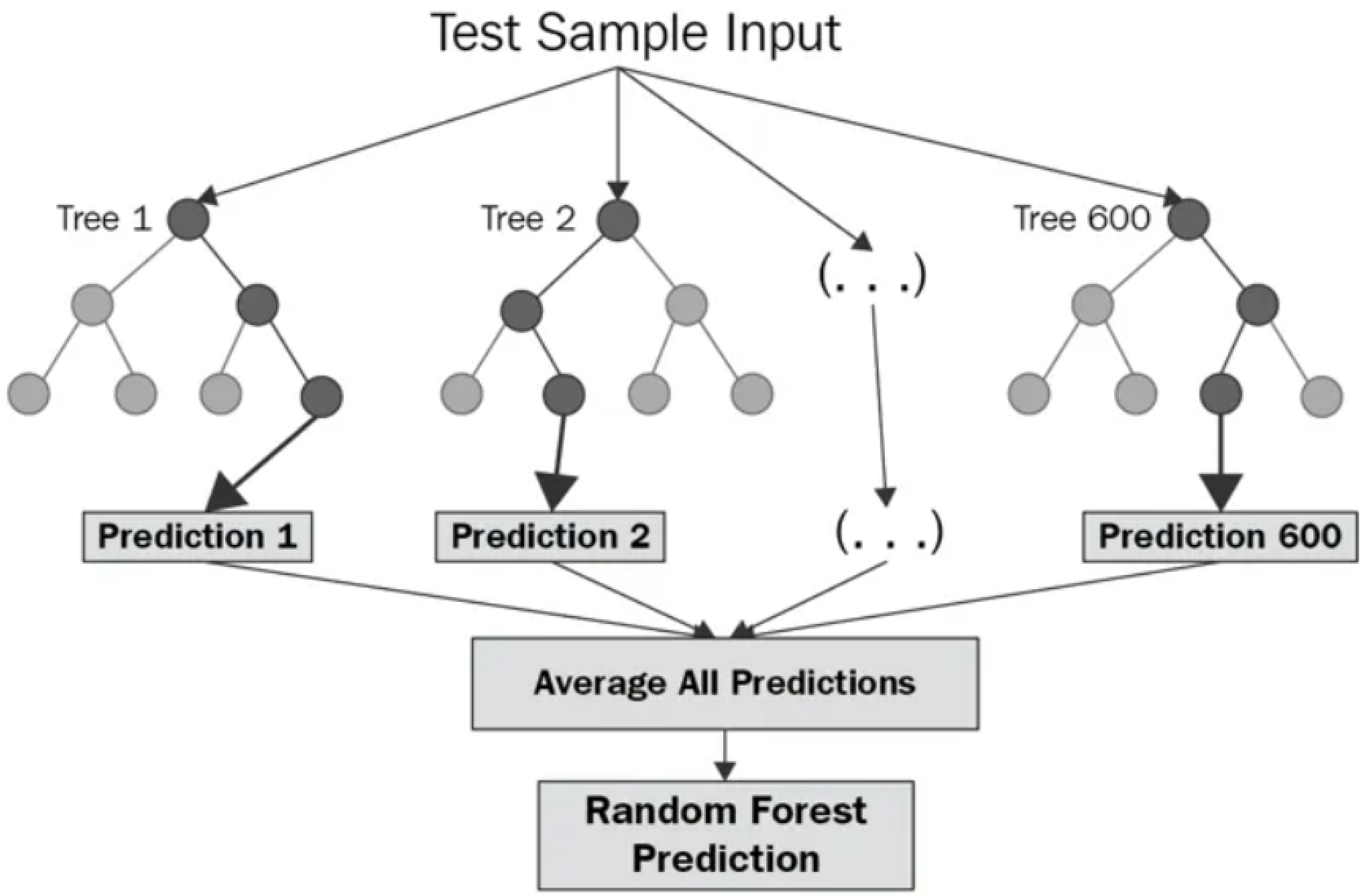
Random Forest Benefits

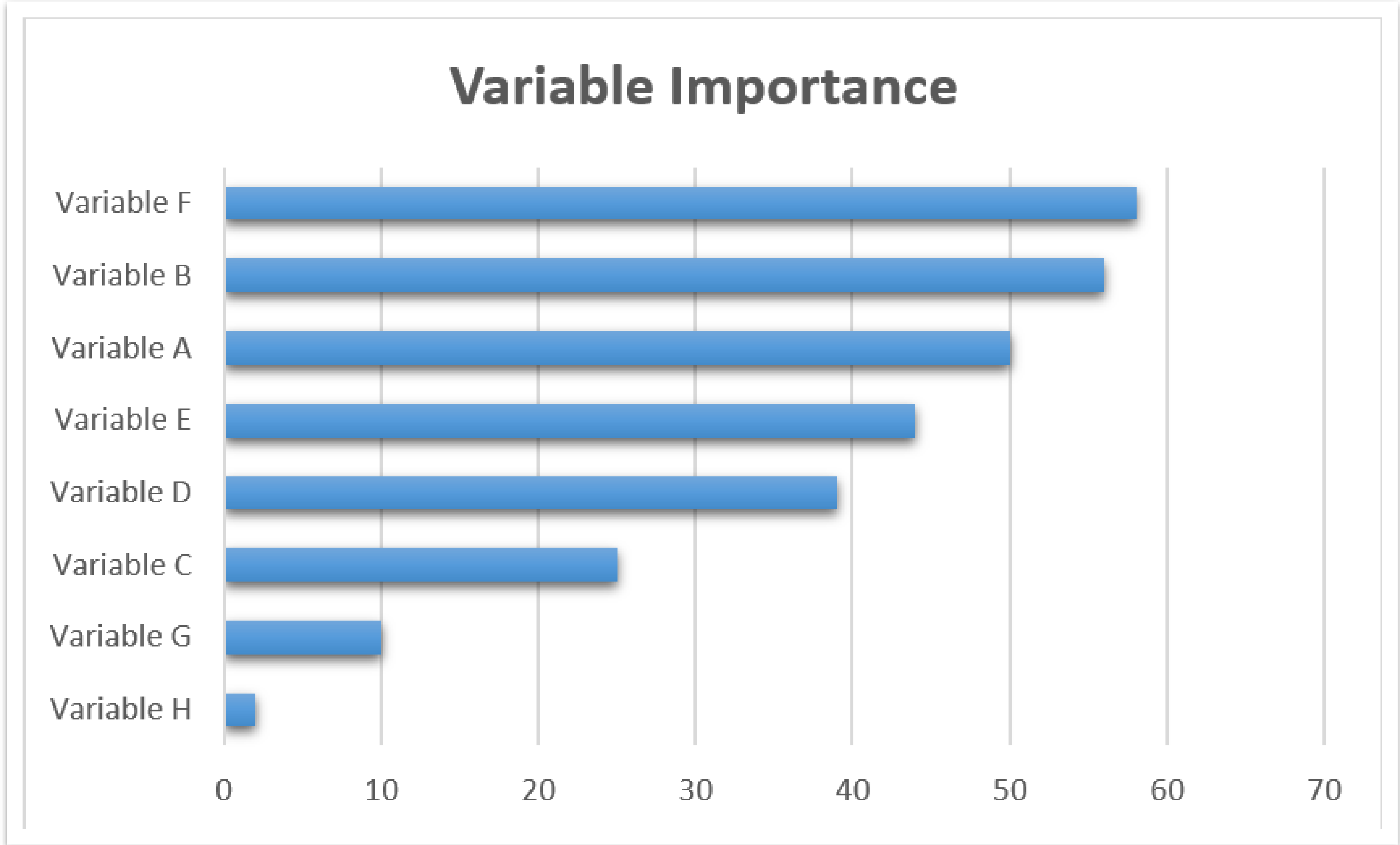
From a single tree to a forest…



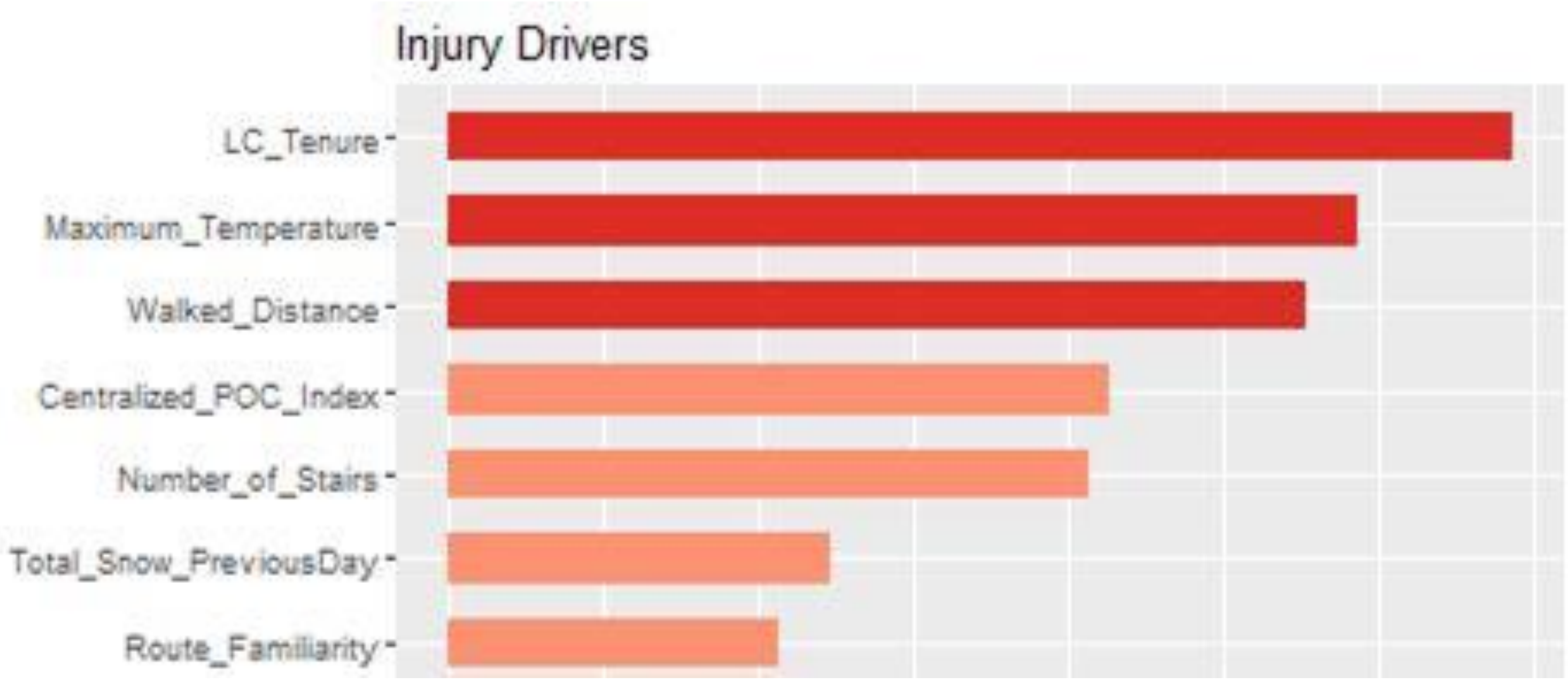


Interpreting Results

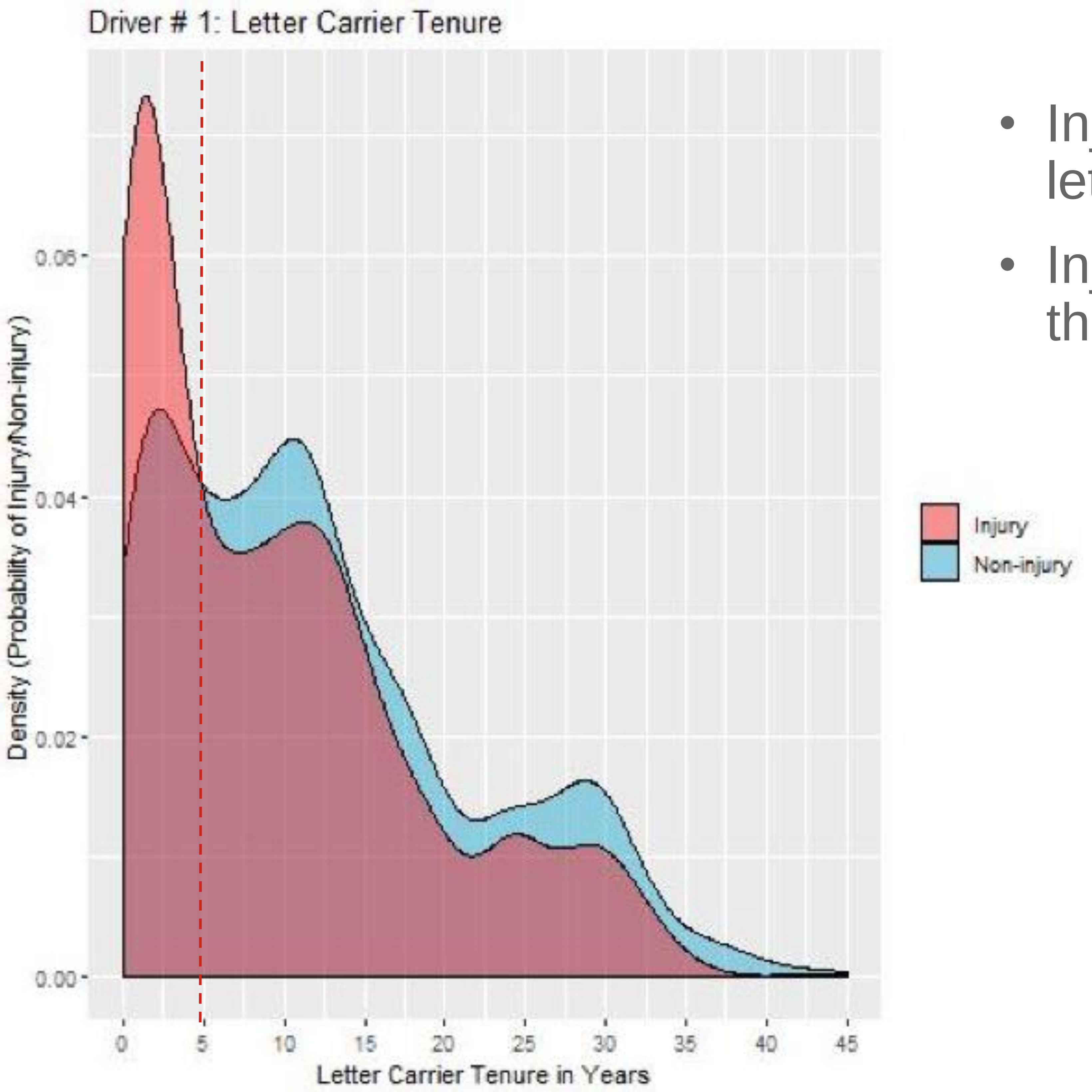




Drivers of Injuries to LCs

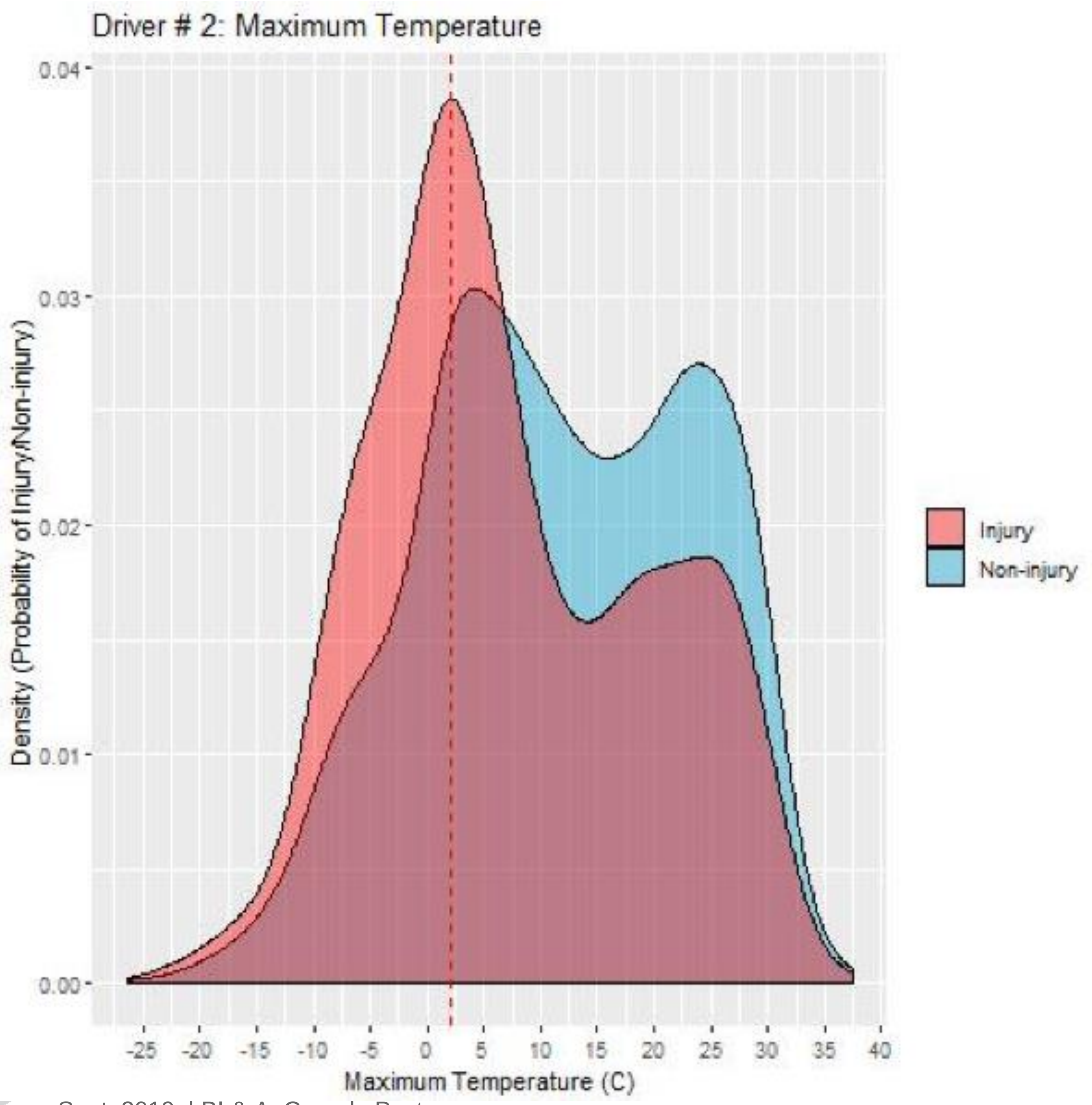


Driver #1: Letter Carrier Tenure



- Injuries most likely during first 2 years as letter carrier
- Injuries still over-indexed up to 5 years into the job

Driver #2: Maximum Daily Temperature



- Injuries most likely to occur when temperature is around freezing
- Days with a high of 2 degrees Celsius have highest threat of injuries

Driver #3: Distance Walked on Route

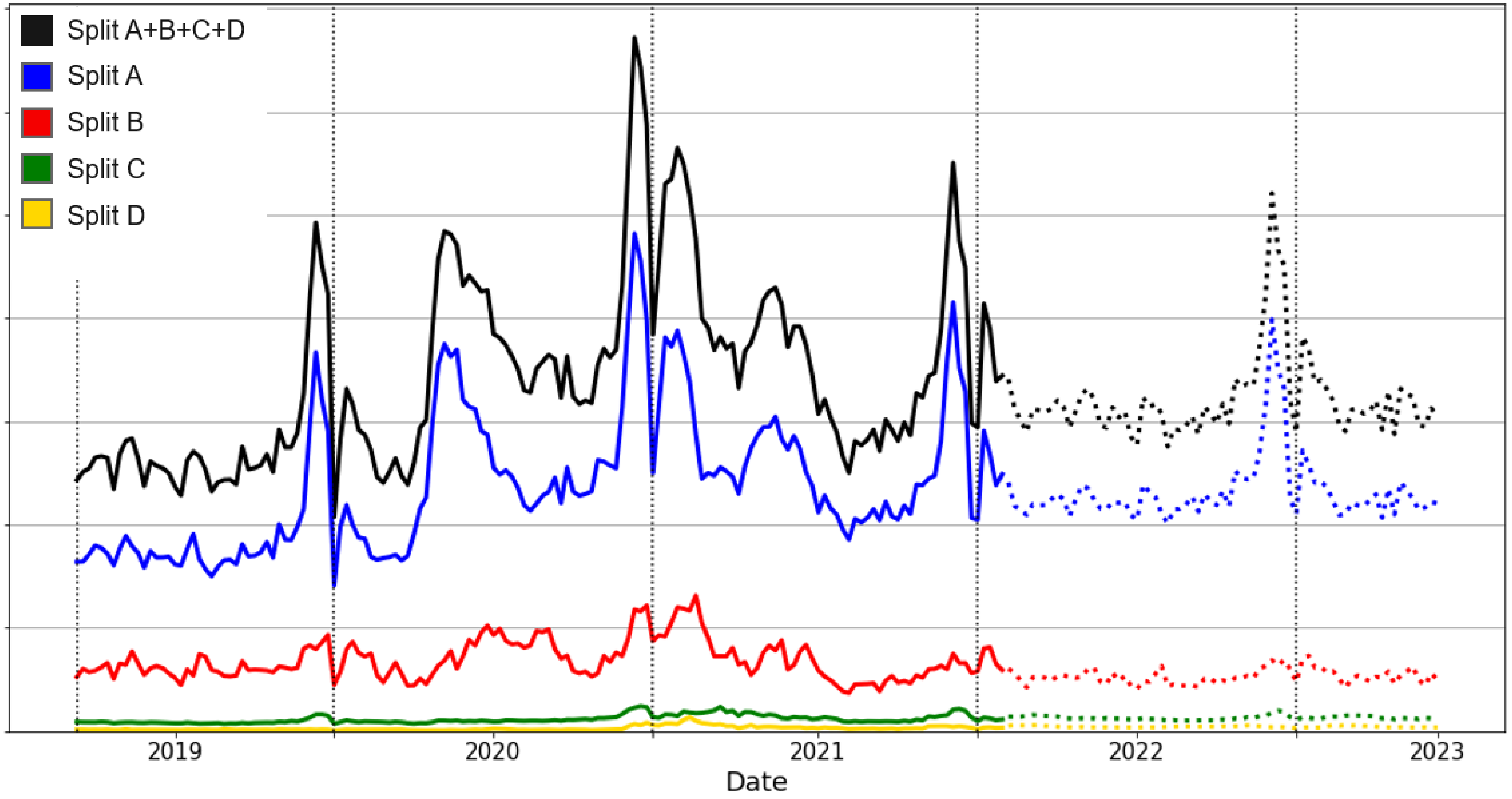
- Longer routes in walked distance (not time) increase risk of injuries
- The tipping point for route length is about 25K feet.

Predictive Analytics

What will happen?

Predictive Analytics - Forecasting

- Predicting future parcel volume nationally and by plant
- *Based on past volume seasonal trends & business insights*
- *Time-Series Model*
- *Seasonality, day of the week*

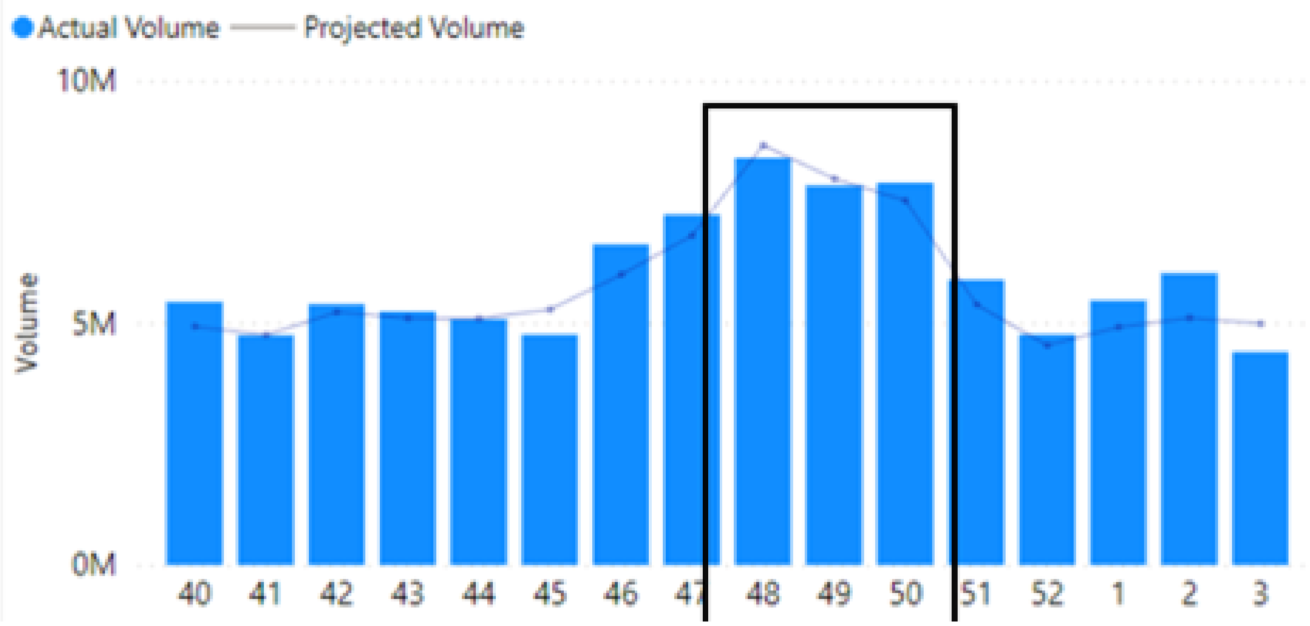


Evaluating Performance

Nationally – Within 1.6% of actual

Plant Level – Within 10%

Plants with > 10% difference are investigated further



Plant	% Difference
NATIONAL	1.6%
TORONTO GATEWAY	0.2%
VANCOUVER PPC	-1.8%
LEO BLANCHETTE MPP	12.8%
KITCHENER MPP	41.6%
QUEBEC MPP	9.2%
ST. JOHNS MPP	9.0%
TORONTO YDC	8.8%
MONCTON MPP	8.5%
SASKATOON MPP	7.5%
OTTAWA MPP	6.6%
THUNDER BAY MPP	6.1%
EDMONTON MPP	5.3%
SAINT JOHN MPP	3.5%
CALGARY MPP	1.5%
SUDBURY MPP	-2.4%
HALIFAX MPP	-3.3%
REGINA MPP	-4.2%
VICTORIA MPP	-7.3%
WINNIPEG MPP	-7.4%
LONDON MPP	-8.6%

Prescriptive Analytics

How can we make things happen?

Prescriptive Analytics - Simulations

- Peak Season Plant Management
- Which customers can we sell more volume to?
- Induction plant

	564,885		564,885		564,885		564,885		564,885		564,885		564,885
	56,455		56,455		56,455		56,455		56,455		56,455		56,455
	508,430		508,430		508,430		508,430		508,430		508,430		508,430
MONCTON MPP		OTTAWA MPP		QUEBEC MPP		REGINA MPP		SAINT JOHN MPP		SASKATOON MPP		ST. JOHNS MPP	
Baseline Forecast	Sales Input	Baseline Forecast	Sales Input	Baseline Forecast	Sales Input	Baseline Forecast	Sales Input	Baseline Forecast	Sales Input	Baseline Forecast	Sales Input	Baseline Forecast	Sales Input
16,894	16,894	16,894	16,894	16,894	16,894	16,894	16,894	16,894	16,894	16,894	16,894	16,894	16,894
435	435	435	435	435	435	435	435	435	435	471	471	307	307
994	1,719	994	1,719	994	1,719	376	1,719	994	1,719	406	796	241	471
18,144	18,144	18,144	18,144	18,144	18,144	481	18,144	18,144	18,144	700	700	102	102
4,136	4,173	4,136	4,173	4,136	4,173	3,521	4,173	4,136	4,173	1,686	1,701	1,952	1,969
768	693	768	693	768	693	1,896	693	768	693	422	301	66	47
1,543	1,625	1,543	1,625	1,543	1,625	405	1,625	1,543	1,625	442	480	297	324
275	275	275	275	275	275	532	275	275	275	37	37	21	21
1,917	1,612	1,917	1,612	1,917	1,612	1,976	1,612	1,917	1,612	786	661	1,408	1,184
468	468	468	468	468	468	973	468	468	468	86	86	20	20
2,037	2,037	2,037	2,037	2,037	2,037	3,226	2,037	2,037	2,037	1,295	1,295	1,003	1,003
3,135	2,560	3,135	2,560	3,135	2,560	3,785	2,560	3,135	2,560	1,089	829	1,121	853
344	262	344	262	344	262	88	262	344	262	92	37	105	43
1,338	1,338	1,338	1,338	1,338	1,338	510	1,338	1,338	1,338	398	398	471	471
976	1,035	976	1,035	976	1,035	186	1,035	976	1,035	61	72	186	221
488	502	488	502	488	502	373	502	488	502	90	96	55	59
633	633	633	633	633	633	287	633	633	633	131	131	117	117
514	434	514	434	514	434	483	434	514	434	246	200	28	22
4,157	4,157	4,157	4,157	4,157	4,157	17	4,157	4,157	4,157	15	15	6	6
389	389	389	389	389	389	30	389	389	389	13	13	29	30
764	770	764	770	764	770	38	770	764	770	34	40	14	16
594	588	594	588	594	588	22	588	594	588	18	15	10	8

Not real data. Data used for illustrative purposes only.

Prescriptive Analytics - Simulations

- Impact on induction and destination plants

MONCTON MPP	OTTAWA MPP	QUEBEC MPP	REGINA MPP	SAINT JOHN MPP	SASKATOON MPP	ST. JOHNS MPP
564,885	564,885	564,885	564,885	564,885	564,885	564,885
56,455	56,455	56,455	56,455	56,455	56,455	56,455
508,430	508,430	508,430	508,430	508,430	508,430	508,430

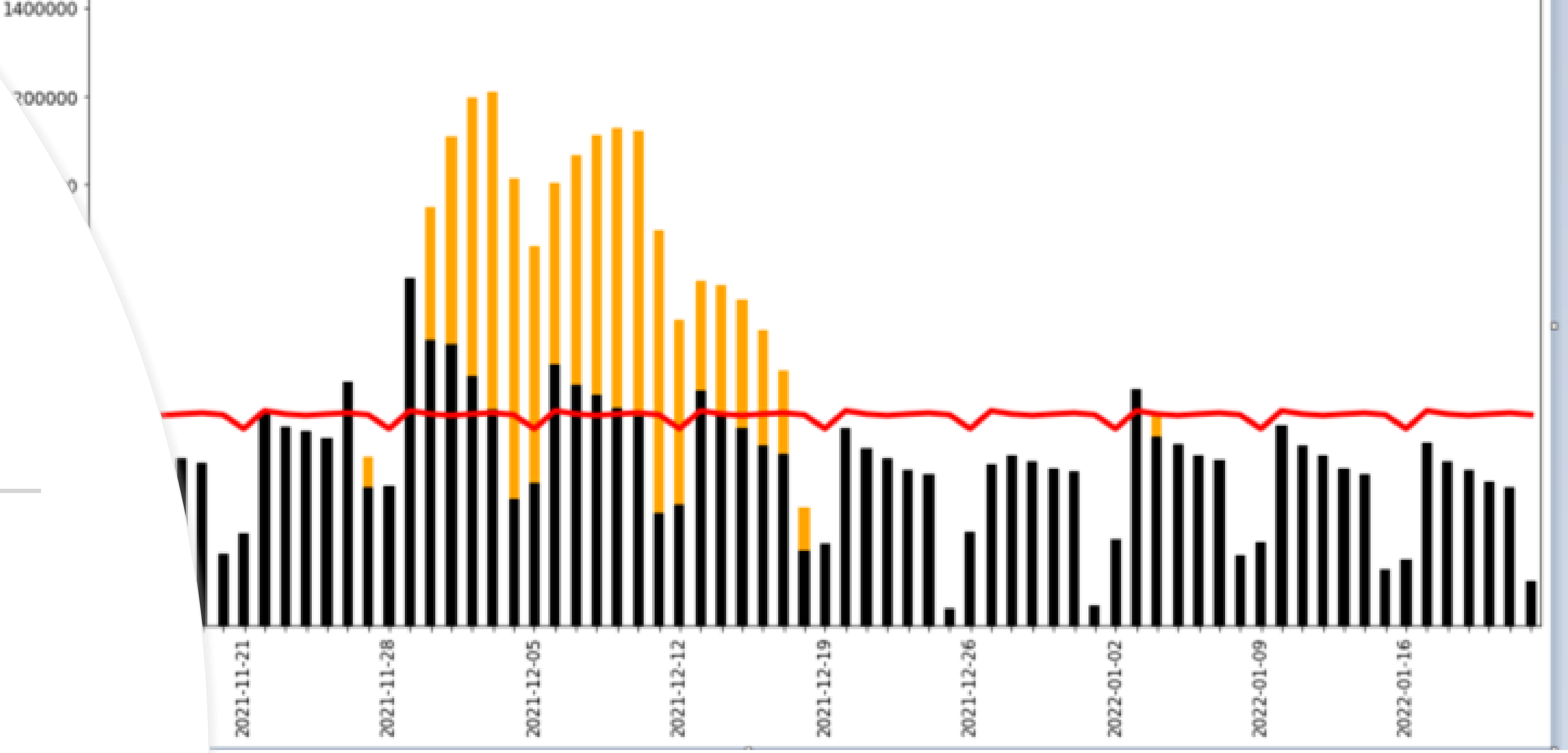
Baseline Forecast	Sales Input
16,894	16,894
435	435
994	1,719
18,144	18,144
4,136	4,173
768	693
1,543	1,625
275	275
1,917	1,612
468	468
2,037	2,037
3,135	2,560
344	262
1,338	1,338
976	1,035
488	502
633	633
514	434
4,157	4,157
389	389
764	770
594	588

4.3%	11.6%	0.5%	0.0%	0.8%	0.4%	0.1%	0.0%
0.2%	30.9%	0.2%	0.3%	0.2%	0.2%	0.1%	0.0%
0.2%	6.2%	0.2%	7.3%	0.7%	0.2%	0.2%	0.0%
0.6%	45.2%	0.9%	1.9%	5.7%	3.5%	1.7%	0.4%
0.4%	12.5%	0.4%	0.1%	2.4%	0.6%	0.9%	0.0%
0.1%	34.3%	0.2%	0.5%	0.3%	0.4%	0.2%	0.1%
0.2%	12.1%	0.2%	0.1%	0.6%	0.0%	0.5%	0.0%
0.2%	18.0%	0.2%	1.9%	1.3%	1.5%	2.0%	0.7%
0.3%	18.0%	0.2%	0.1%	0.4%	0.0%	1.0%	0.0%
1.2%	6.4%	1.3%	1.9%	4.7%	0.7%	3.6%	1.0%
1.1%	1.4%	1.1%	2.8%	2.4%	0.1%	4.4%	0.5%
0.1%	1.6%	0.2%	0.2%	0.3%	0.2%	0.1%	0.0%
0.9%	33.6%	0.9%	1.2%	4.5%	4.2%	0.9%	0.0%
0.0%	35.4%	0.4%	0.6%	6.4%	0.3%	0.4%	0.2%
0.8%	7.5%	0.2%	0.5%	1.0%	0.0%	0.9%	0.1%
0.1%	3.8%	0.2%	0.6%	1.0%	0.6%	0.7%	0.1%
0.3%	74.3%	0.5%	1.0%	1.7%	1.3%	1.1%	0.0%
0.1%	9.1%	0.1%	8.2%	0.1%	0.1%	0.0%	0.0%
0.1%	0.6%	0.1%	0.1%	0.3%	4.3%	0.1%	0.0%
0.1%	18.7%	0.1%	0.1%	0.3%	0.1%	0.1%	0.0%
0.2%	12.1%	0.4%	0.1%	2.9%	2.0%	0.1%	0.0%
0.1%	0.3%	0.1%	0.2%	0.1%	0.1%	0.1%	0.0%
0.1%	98.4%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%
0.1%	0.2%	0.0%	0.1%	0.1%	0.0%	0.0%	0.0%
0.2%	4.5%	0.2%	0.3%	1.1%	0.1%	0.4%	0.1%
0.2%	20.0%	0.2%	0.3%	0.2%	0.0%	0.1%	0.0%
0.0%	8.6%	0.2%	0.6%	1.0%	0.8%	0.2%	0.2%

Not real data. Data used for illustrative purposes only.

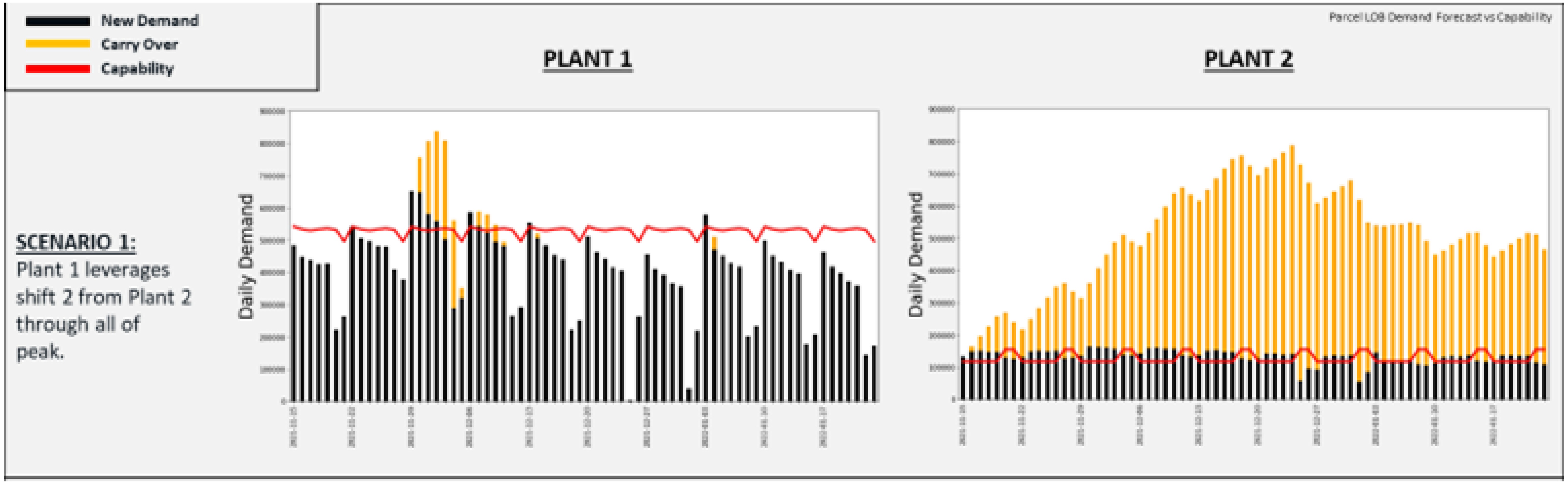
Prescriptive Analytics - Scenarios

- Plant 1 – MAJOR plant with high volume
- # of backlog days
- When will backlog clear?



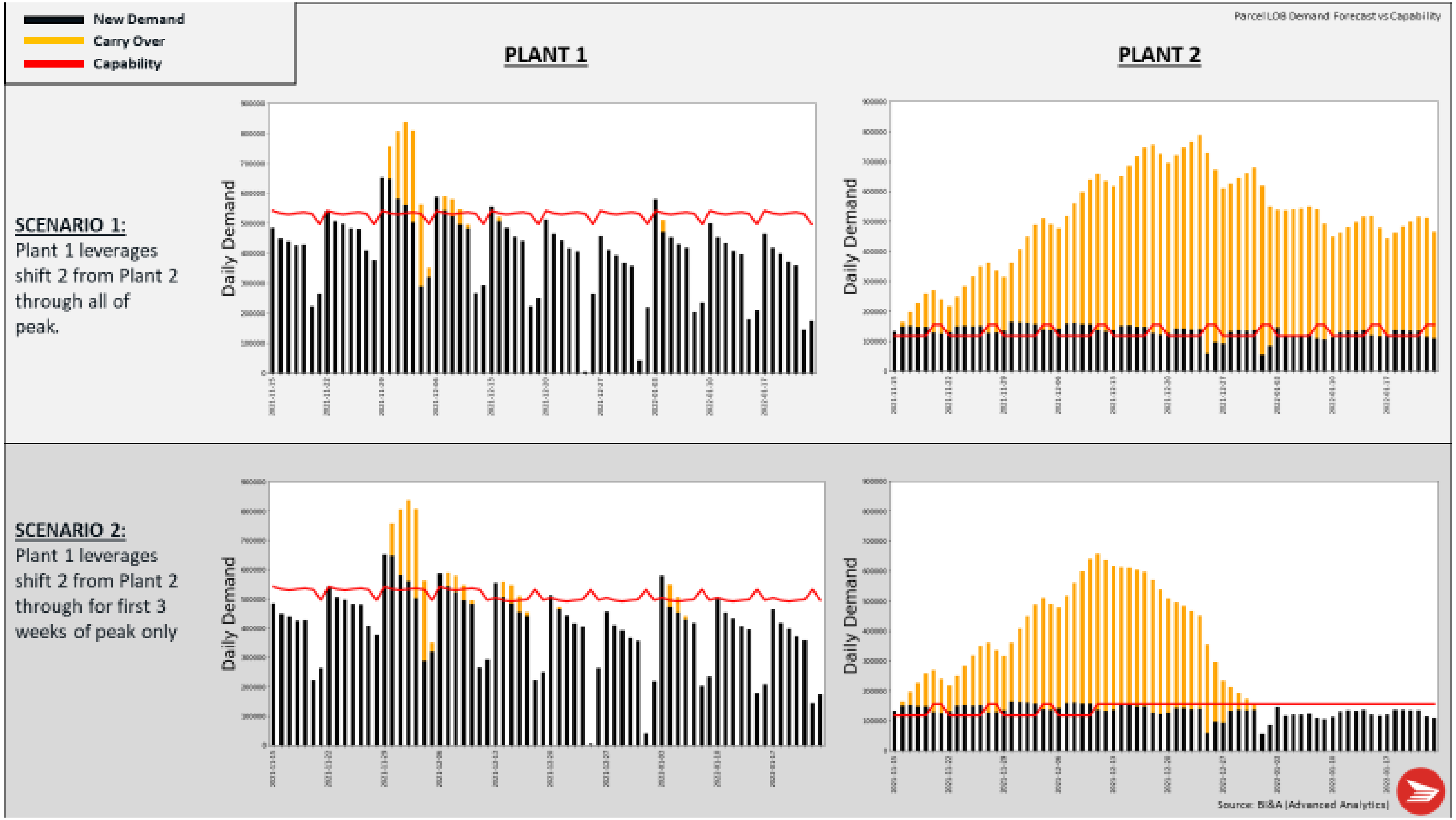
Prescriptive Analytics - Scenarios

- What if we borrow workers from a nearby minor plant to assist in major plant's backlog during peak season?

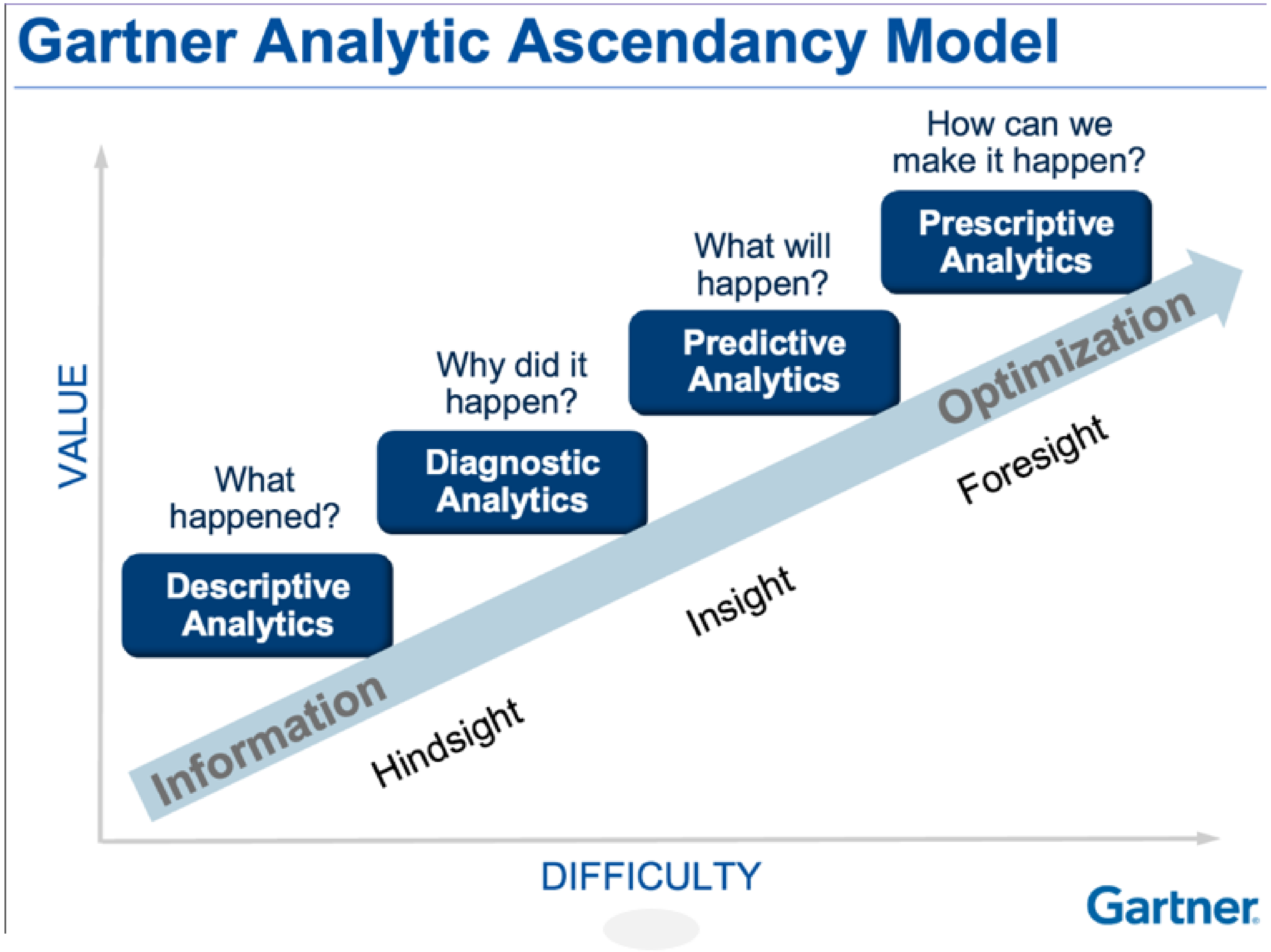


Prescriptive Analytics - Scenarios

- What if we leverage shift 2 from Plant 2 for just first 3 weeks of peak?



What's Beyond Prescriptive?



Integrated Analytics

Integrated Analytics

- Beyond Prescriptive Analytics
- Business friendly apps to aid decision making
- Machine learning and AI utilized without need to request
- Anticipated in 2024

The Evolution of Analytics at Canada Post:

Going Beyond Prediction

Carol Wilson,

Director of Advanced Analytics

Thank you!