The Evolution of Analytics at Canada Post: Going Beyond Prediction

POSTES

POST

CANADA

CANADA

Carol Wilson,
Director of Advanced Analytics



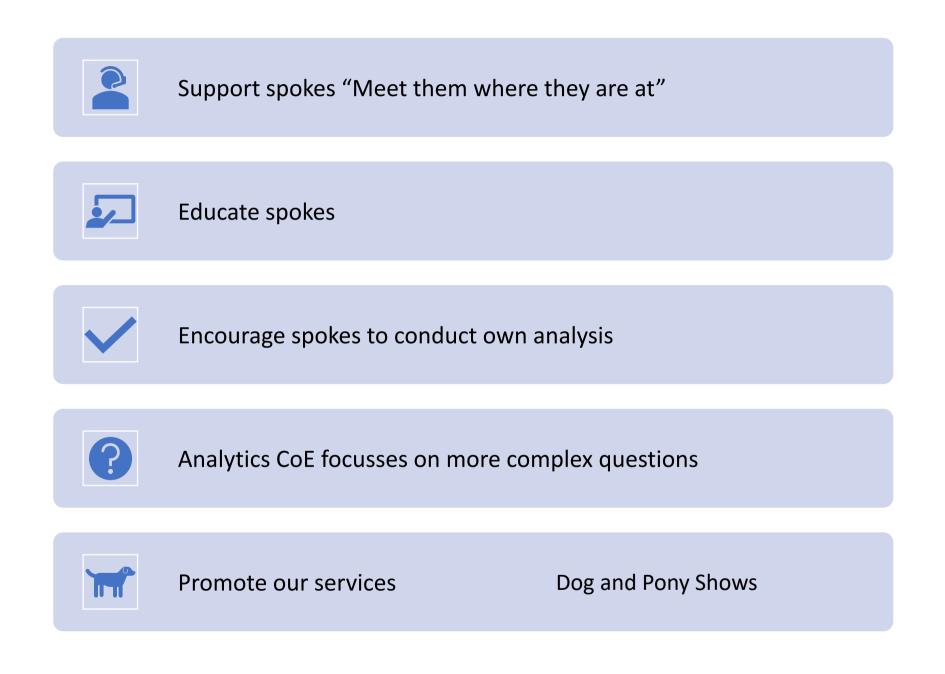
September 21st, 2023

5pm - 9pm

Ottawa | Lansdowne Horticulture Building

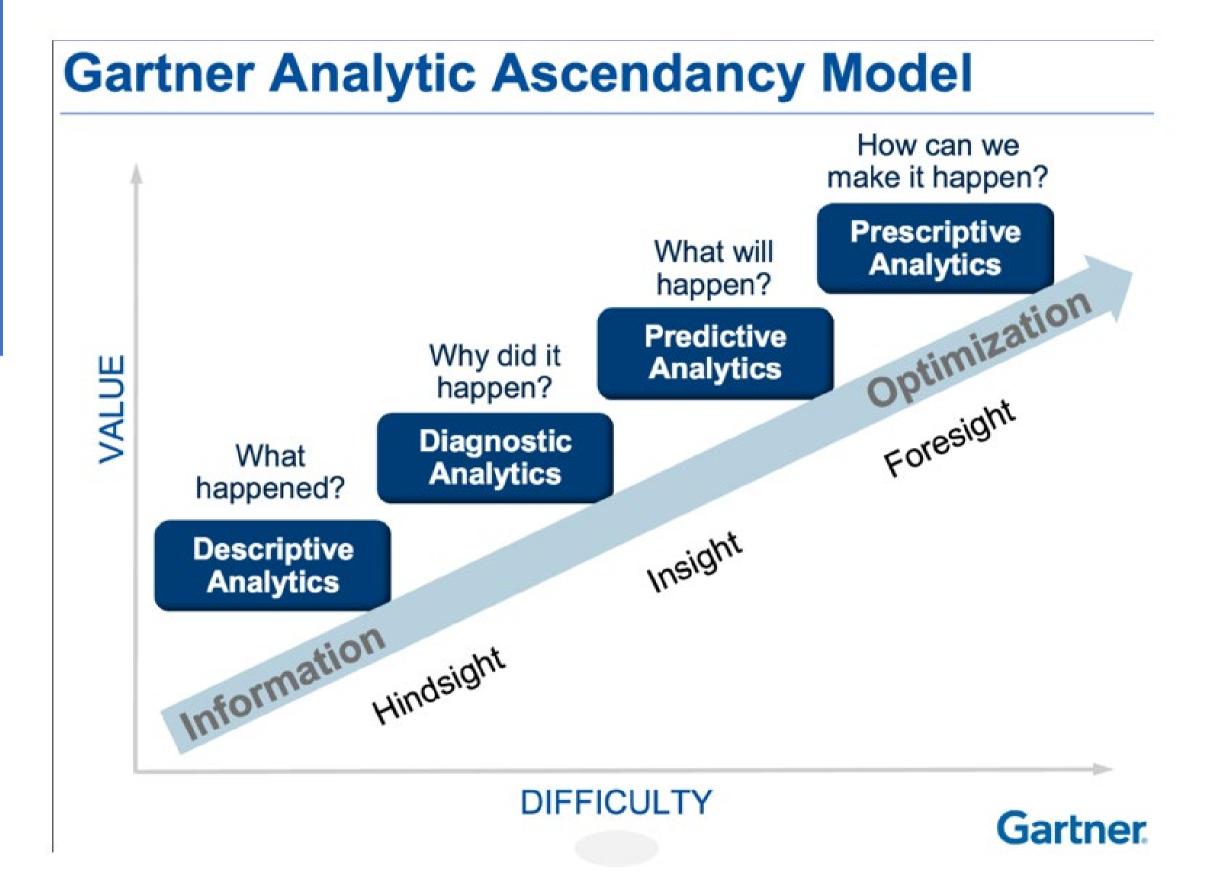
IN PARTNERSHIP WITH

Hub and Spoke Model for Analytics

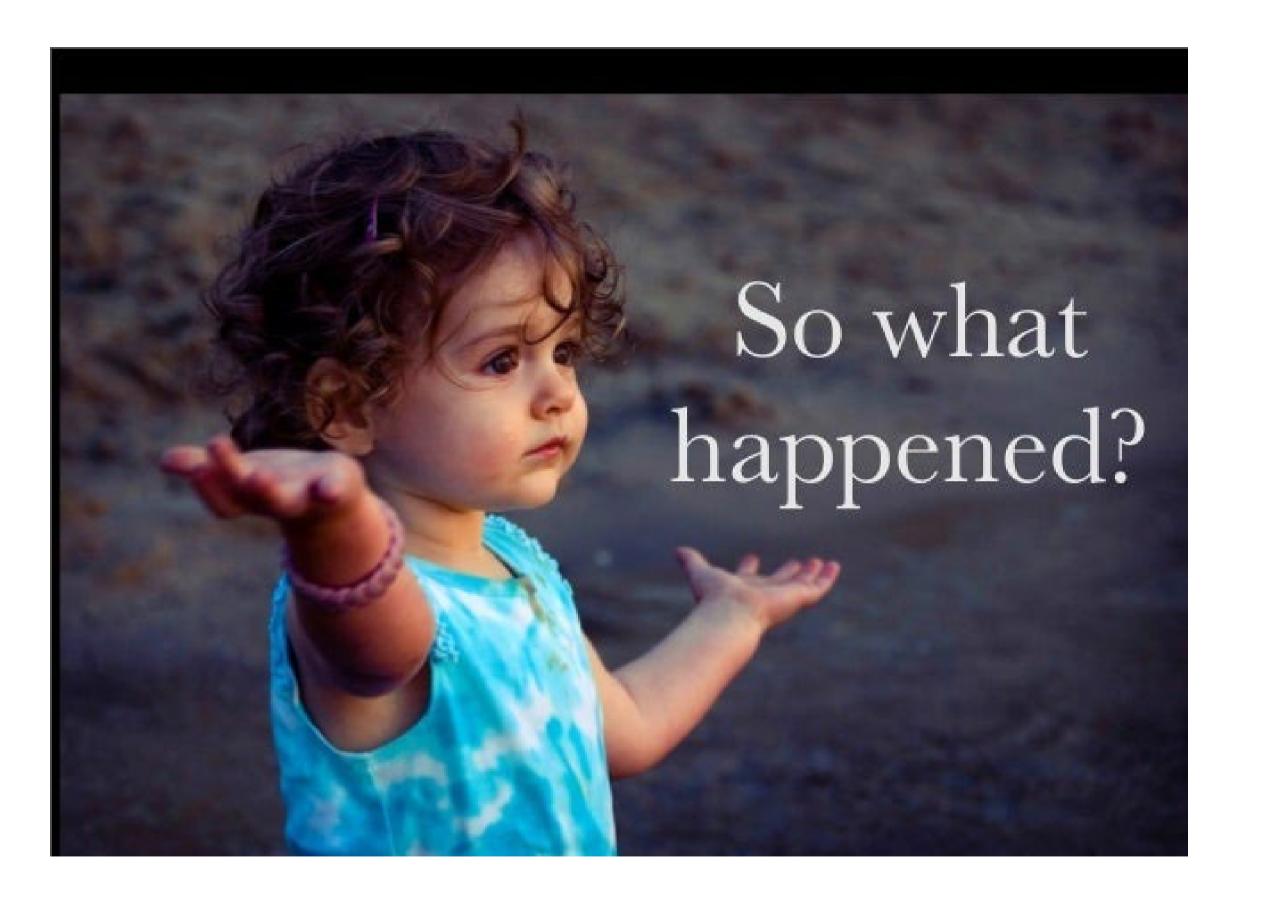




The Evolution of Analytics at Canada Post:
Going Beyond Prediction



Descriptive Analytics



Data Literacy

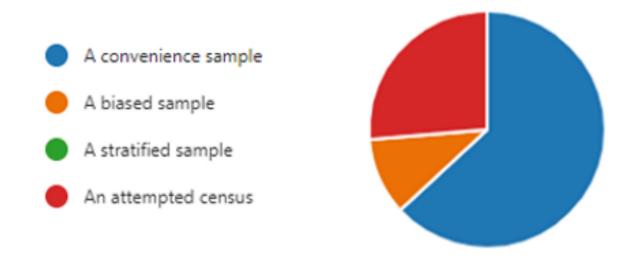
- Need to understand:
 - The type of data you have
 - Census
 - Attempted census
 - Random sample
 - Convenience sample
 - When to use statistical testing



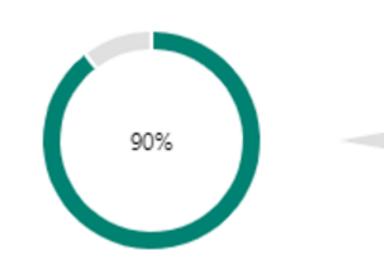
Gap in Sampling Knowledge

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?



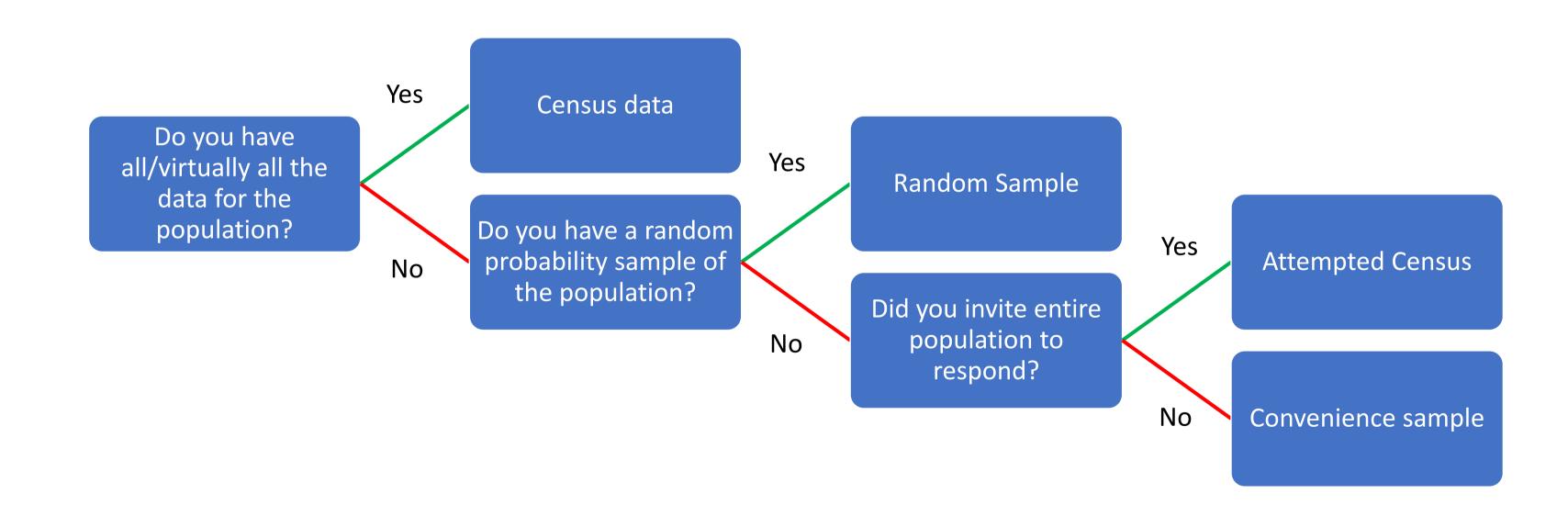
 90% people answered "The margin of error would be much larger than expected" for question 9

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 <u>random</u> <u>probability sample</u>
- With census data we know the data is reliable so the only question is: "Is the difference big enough to care about?"

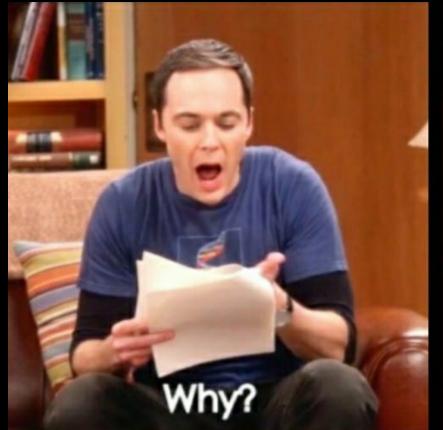


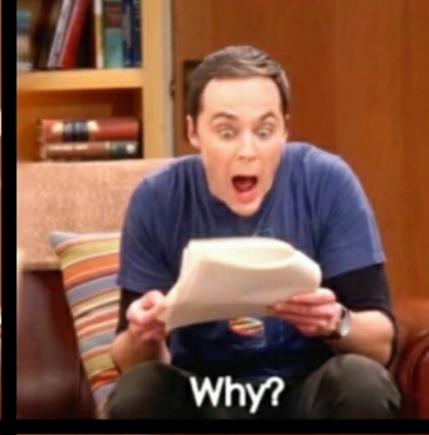
What type of data do I have?

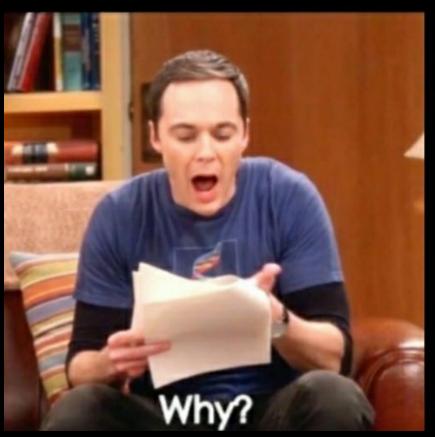


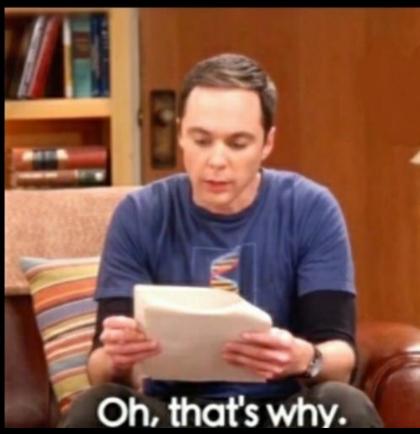
Diagnostic Analytics

Why did it happen?





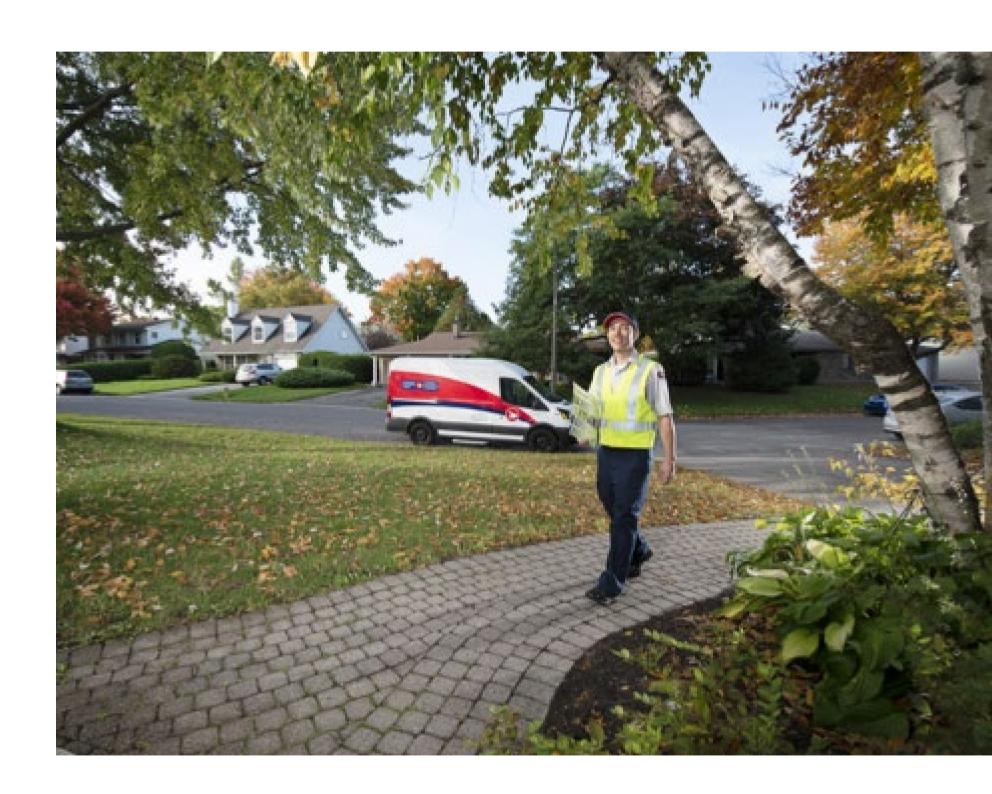




Diagnostic Analytics with Classification Models

- 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

Weather Features

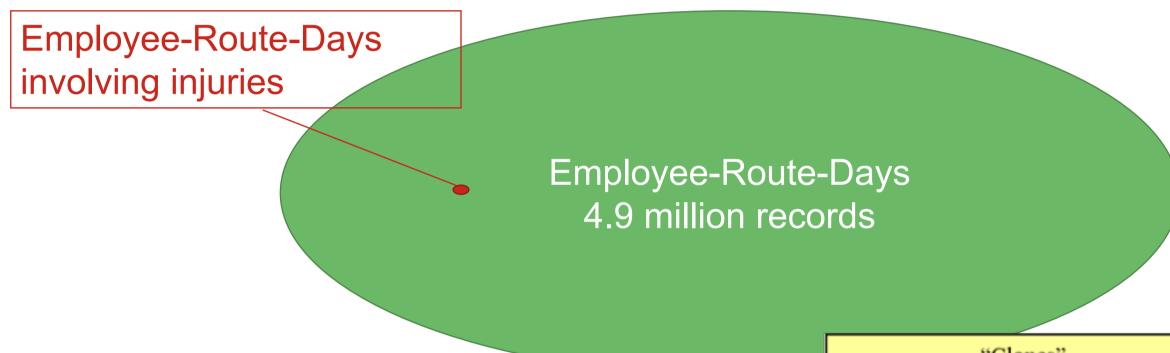
- Maximum temperature for day
- Total of rain for the previous day
- Total of snow for the previous day
- Total rain and snow for delivery day.

Route Characteristic Features

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



Random Forest Model – Design and Samples



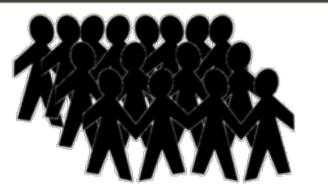
Synthetic Minority Over-Sampling Technique (SMOTE): The SMOTE technique is preferred over *sampling with replacement* because it uses the information available for existing group members to assign reasonable new values to the new sample. Rather than merely replicating existing records in the under-represented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

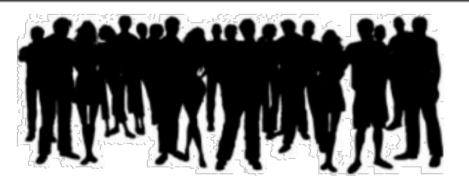
"Clones"

"Offspring"

Sampling with Replacement

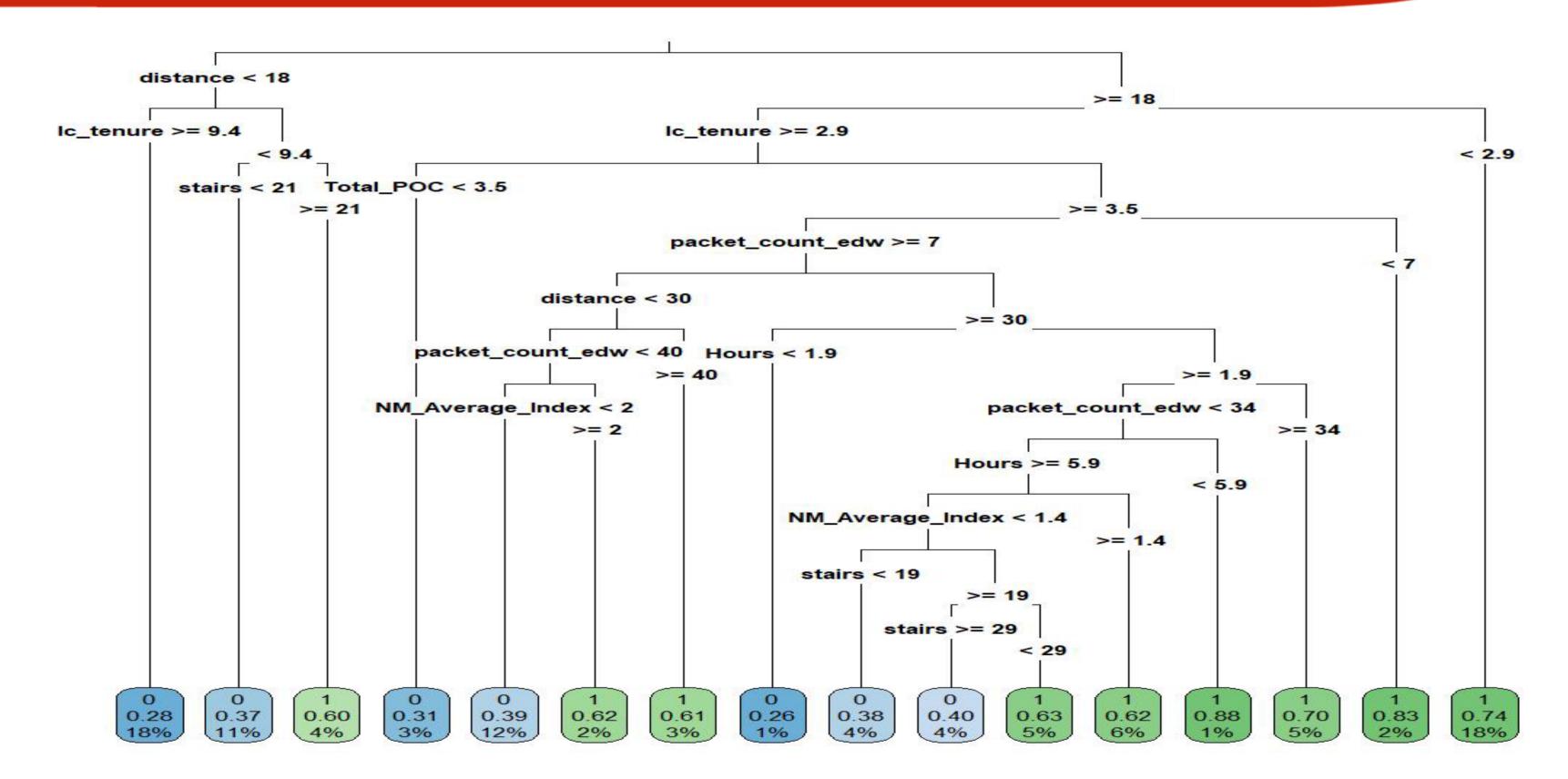
SMOTE





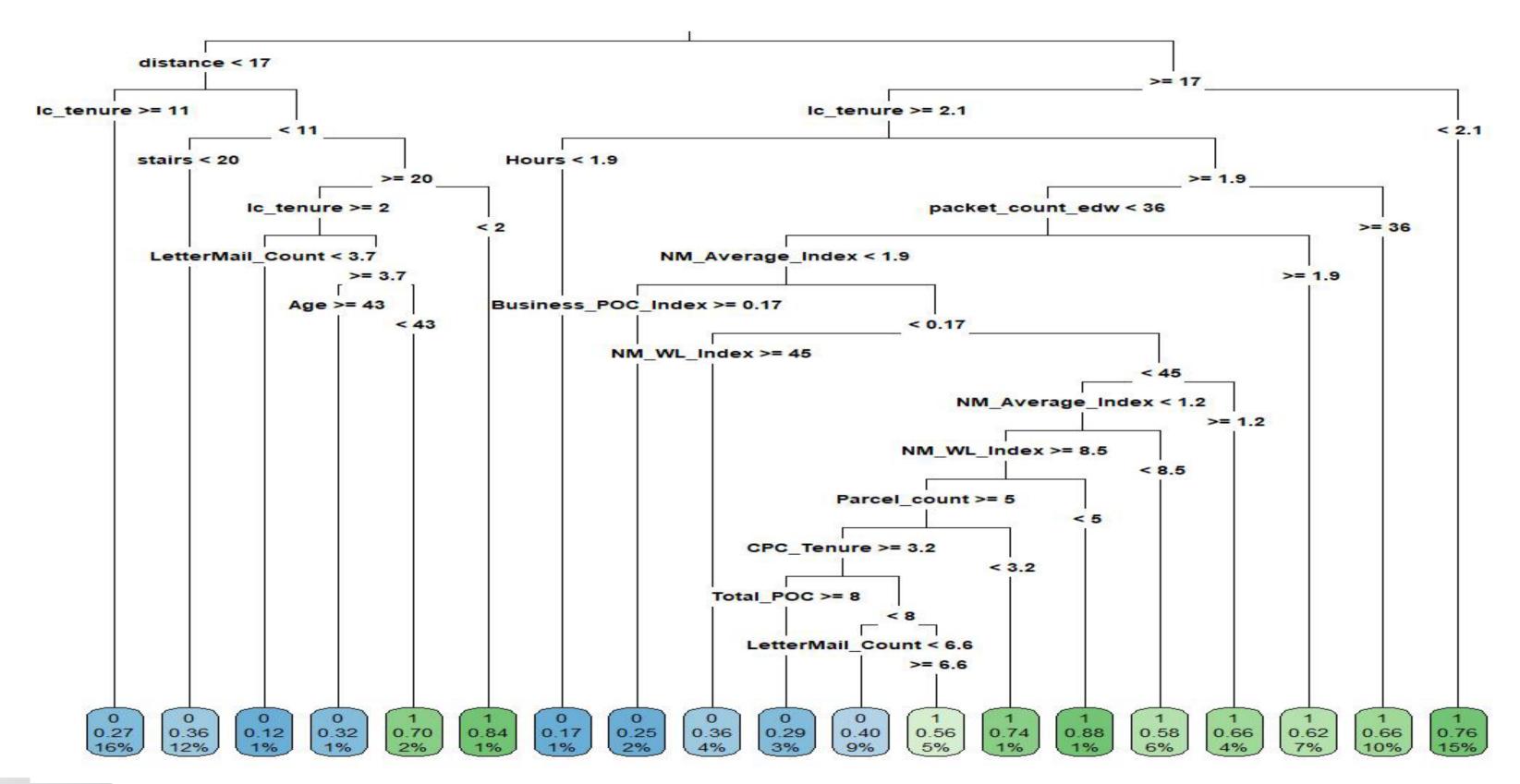


Classic Decision Tree – Random Sample 1





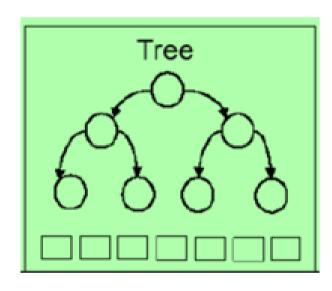
Classic Decision Tree – Random Sample 2



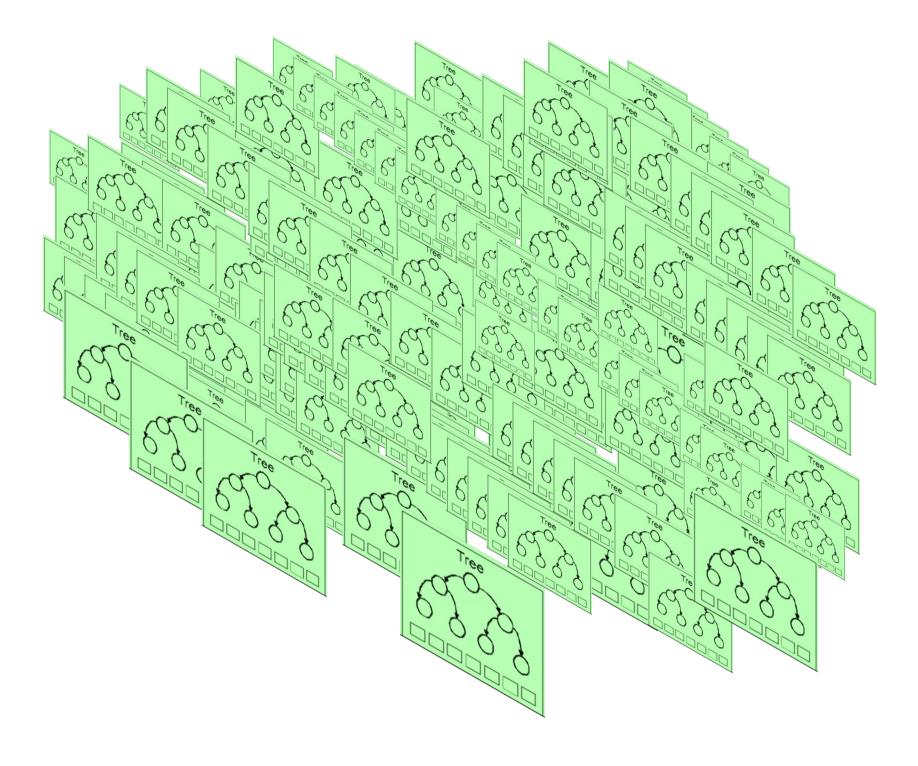


Random Forest Benefits

From a single tree

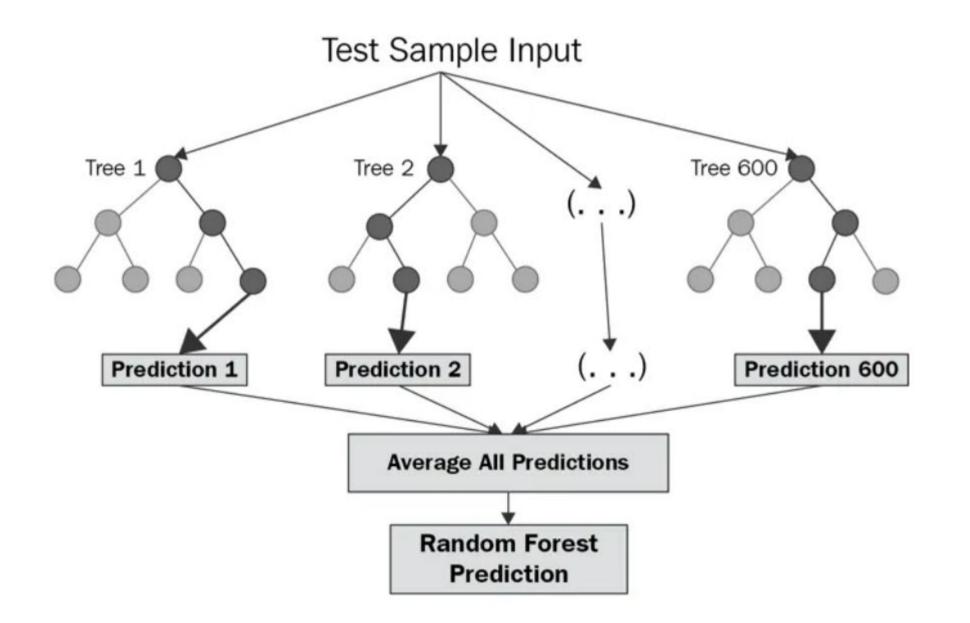


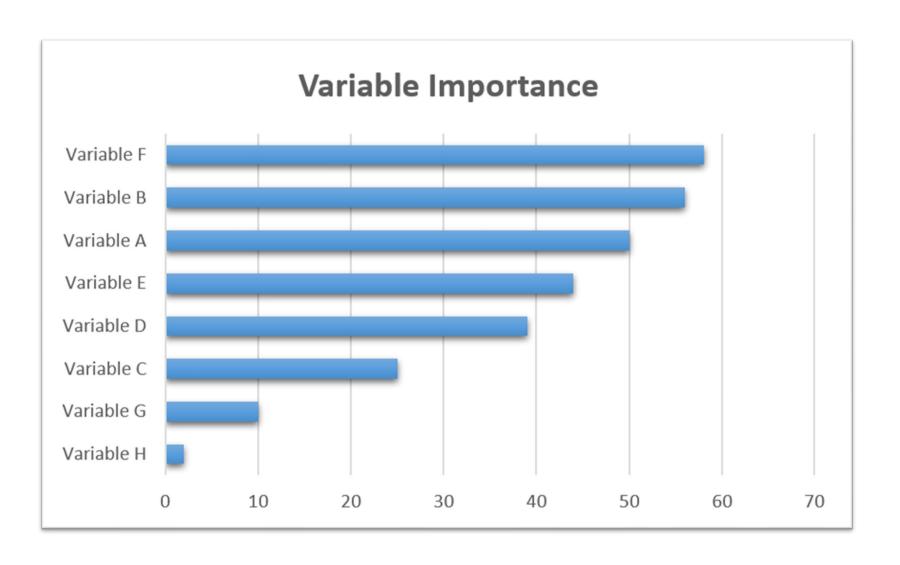
to a forestooo





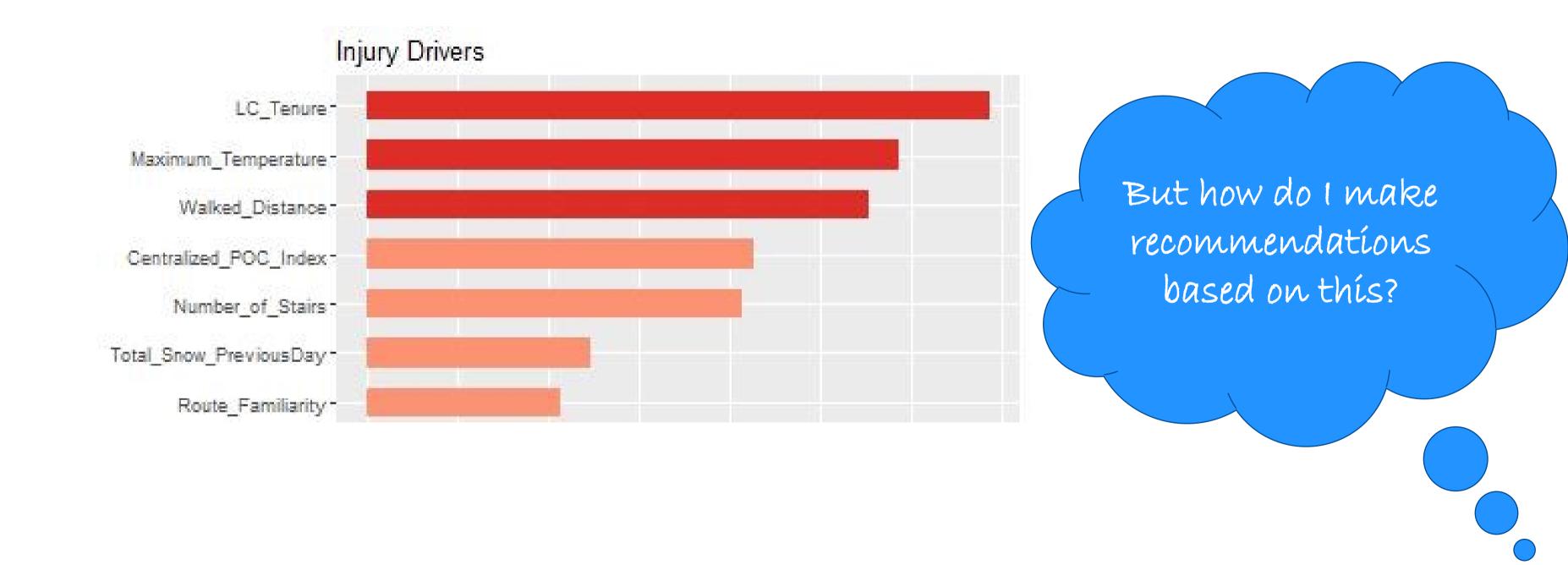
Interpreting Results





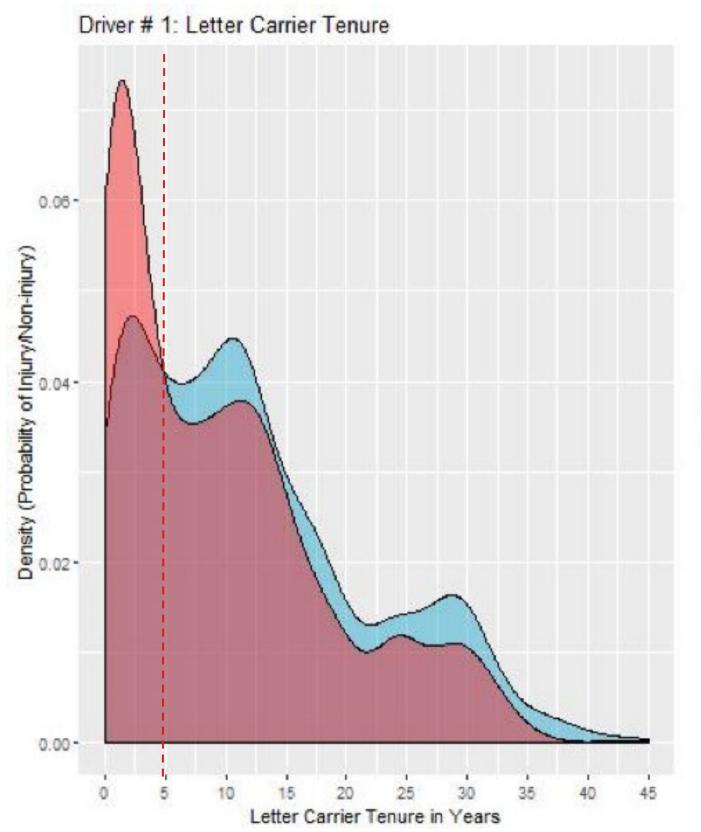


Drivers of Injuries to LCs





Driver #1: Letter Carrier Tenure

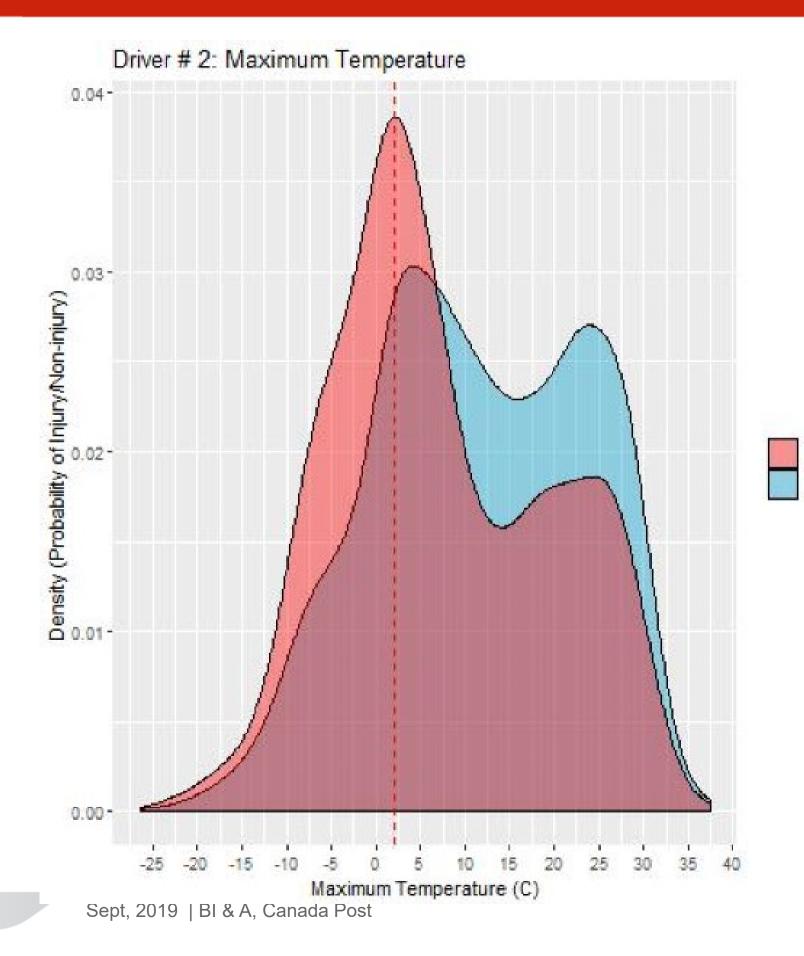


- Injuries most likely during first 2 years as letter carrier
- Injuries still over-indexed up to 5 years into the job





Driver #2: Maximum Daily Temperature



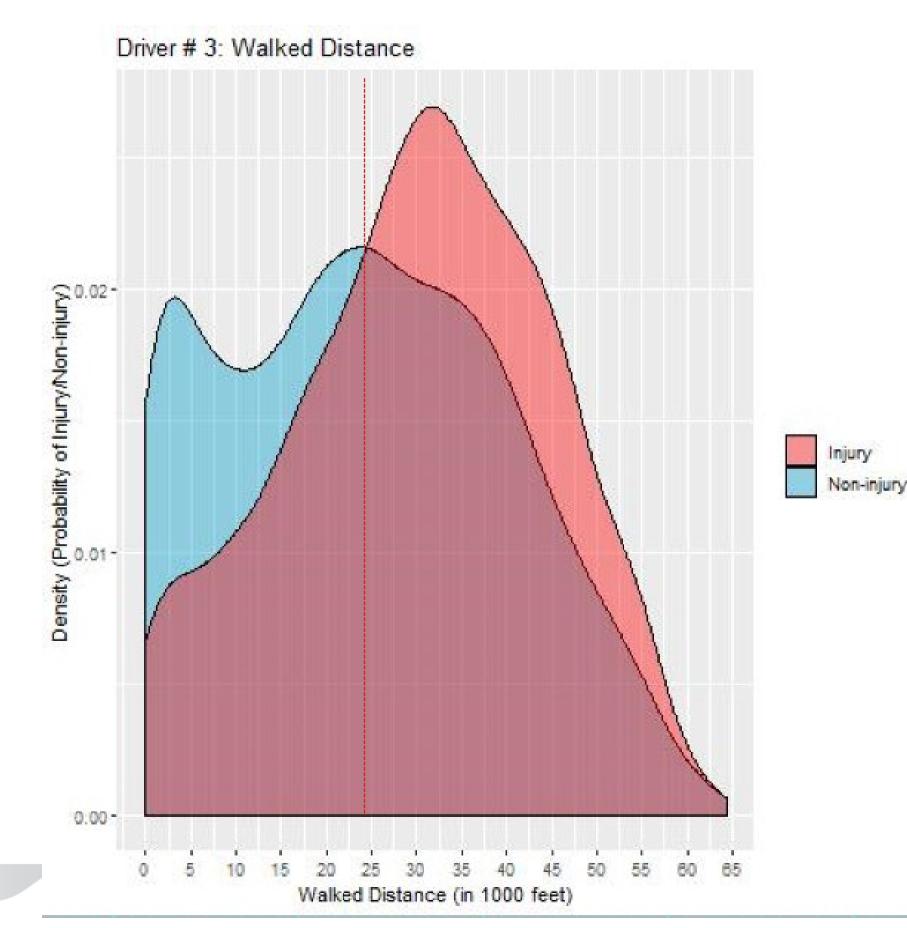
 Injuries most likely to occur when temperature is around freezing

Injury

 Days with a high of 2 degrees Celsius have highest threat of injuries



Driver #3: Distance Walked on Route



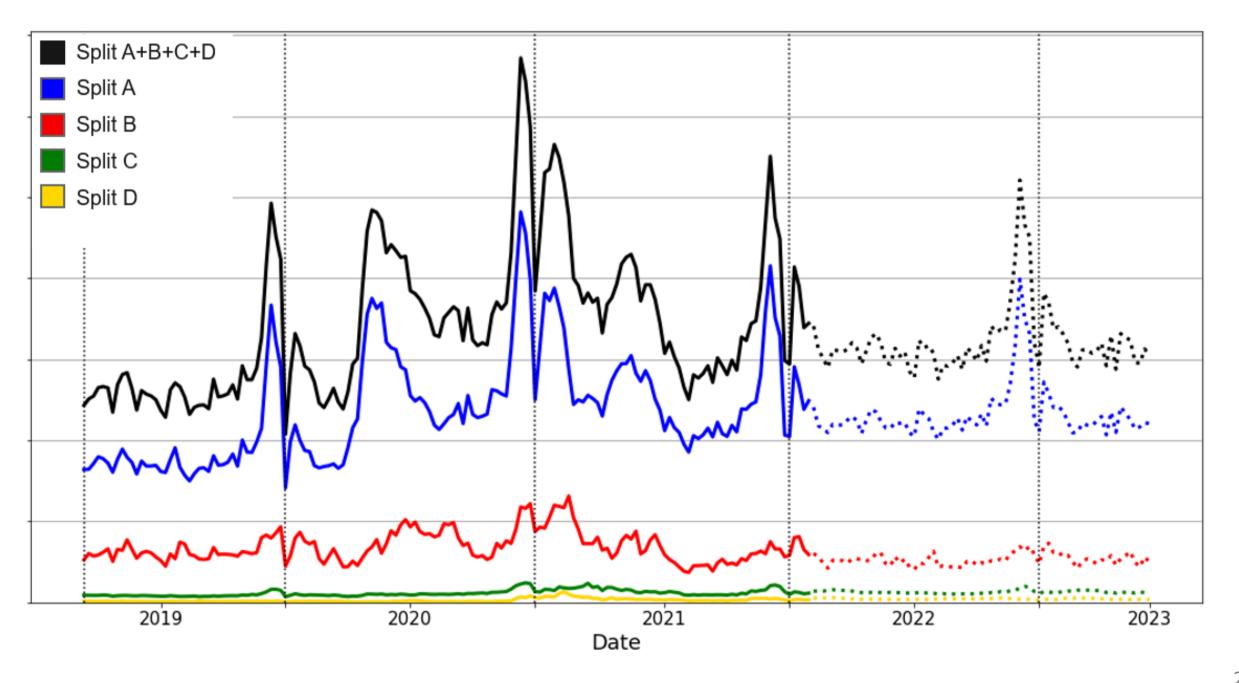
- Longer routes in walked distance (not time) increase risk of injuries
- The tipping point for route length is about 25K feet.





Predictive Analytics 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

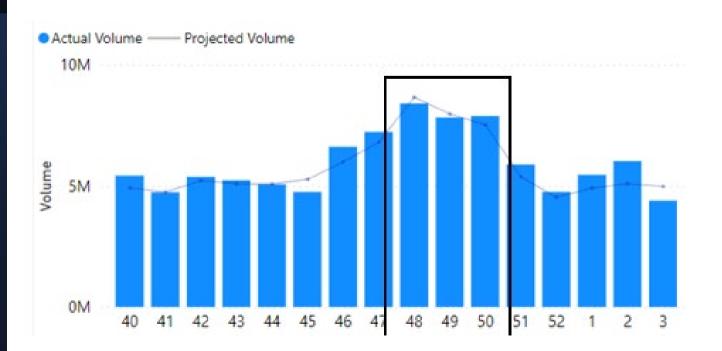


Evaluating Performance

Nationally – Within 1.6% of actual

Plant Level – Within 10%

Plants with > 10% difference are investigated further

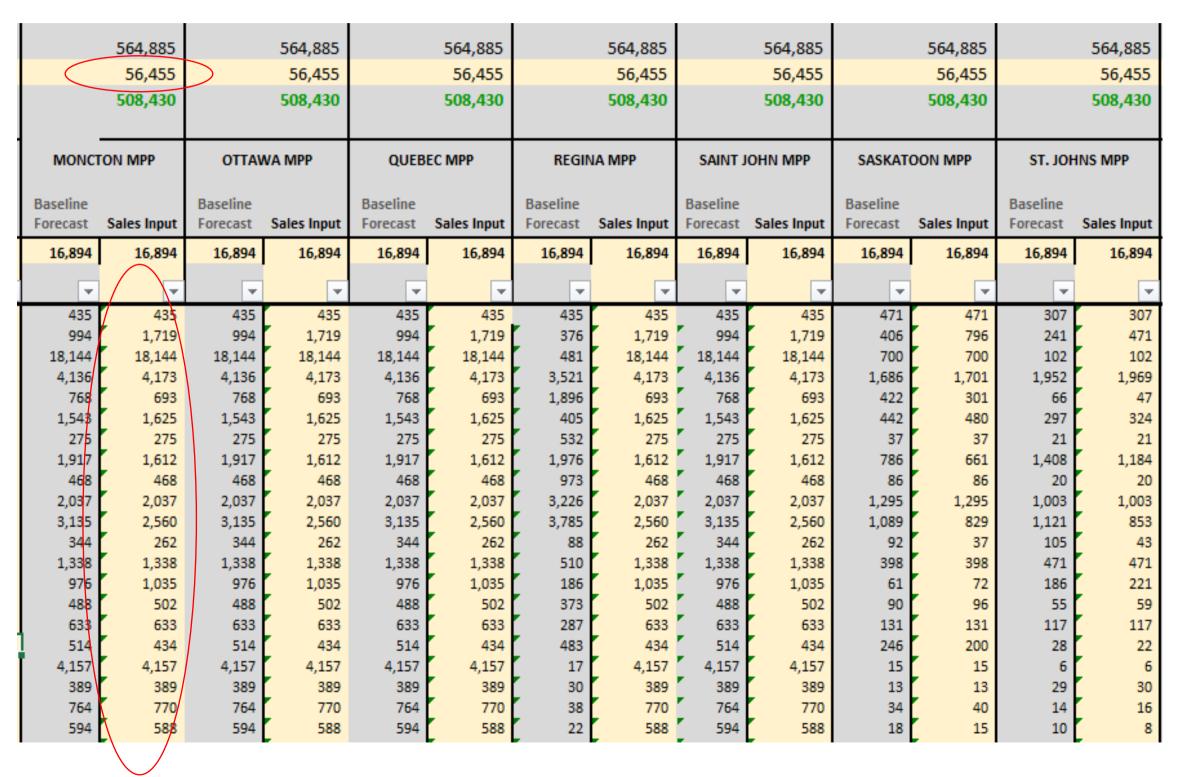


Plant	% Difference				
NATIONAL	1.6%				
TORONTO GATEWAY	0.2%				
VANCOUVER PPC	-1.8%				
LEO BLANCHETTE MPP	12.8%				
KITCHENER MPP	41.6%				
QUEBEC MPP	9.2%				
ST. JOHNS MPP	9.0%				
TORONTO YDC	8.8%				
MONCTON MPP	8.5%				
SASKATOON MPP	7.5%				
OTTAWA MPP	6.6%				
THUNDER BAY MPP	6.1%				
EDMONTON 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



Prescriptive Analytics - Simulations

Impact on induction and destination plants

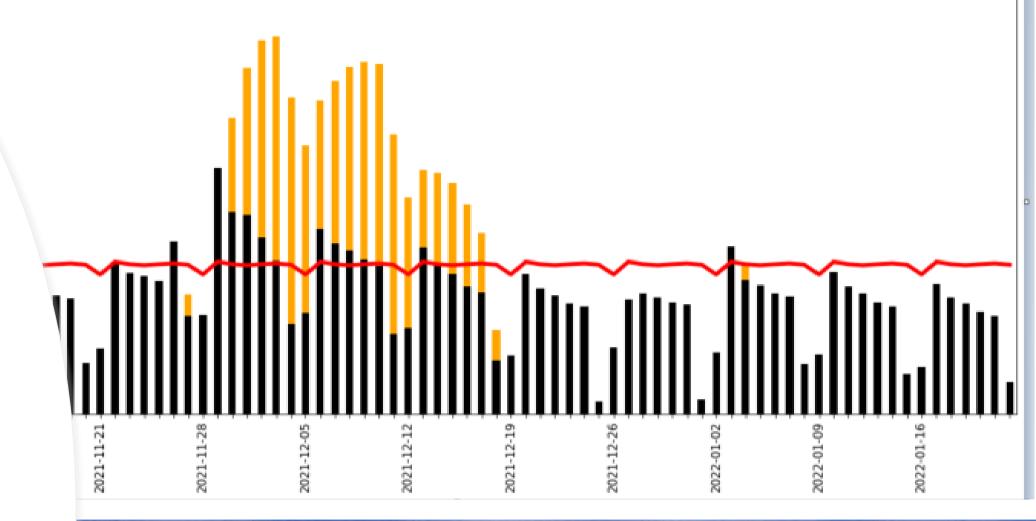
	564,885	564,885	564,885	564,885	564,88	35 564,885	5	564,885	
	56,455	56,455	56,455	56,455	56,45	55 56,455	5	56,455	
508,430		508,430	0 508,430	508,430	508,430	508,430	508,430		
MONCTON MPP		OTTAWA MPP	QUEBEC MPP	REGINA MPP	SAINT JOHN MPP	SASKATOON MPP	P ST. JOHNS MPP		
		4.3%	11.6%	0.5%	0.0%	0.8%	0.4%	0.1%	0.09
Baseline	Calaa laasa	0.2%	30.9%	0.2%	0.3%	0.2%	0.2%	0.1%	0.0
orecast	Sales Input	0.2%	6.2%	0.2%	7.3%	0.7%	0.2%	0.2%	0.0
16,894	16,894	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
435	435	0.1%	34.3%	0.2%	0.5%	0.3%	0.4%	0.2%	0.1
994	1,719	0.2%	12.1%	0.2%	0.1%	0.6%	0.0%	0.5%	0.0
18,144 4,136	18,144 4,173	0.2%	18.0%	0.2%	1.9%	1.3%	1.5%	2.0%	0.7
768	693	0.3%	18.0%	0.2%	0.1%	0.4%	0.0%	1.0%	0.0
1,543	1,625	1.2%	6.4%	1.3%	1.9%	4.7%	0.7%	3.6%	1.0
275	275	1.1%	1.4%	1.1%	2.8%	2.4%	0.1%	4.4%	0.5
1,917 468	1,612 468	0.1%	1.6%	0.2%	0.2%	0.3%	0.2%	0.1%	0.0
2,037	2,037	0.9%	33.6%	0.9%	1.2%	4.5%	4.2%	0.9%	0.0
3,135	2,560	0.0%	35.4%	0.4%	0.6%	6.4%	0.3%	0.4%	0.2
344	262	0.8%	7.5%	0.2%	0.5%	1.0%	0.0%	0.9%	0.1
1,338 976	1,338 1,035	0.1%	3.8%	0.2%	0.6%	1.0%	0.6%	0.7%	0.1
488	502	0.3%	74.3%	0.5%	1.0%	1.7%	1.3%	1.1%	0.0
633	633	0.1%	9.1%	0.1%	8.2%	0.1%	0.1%		
514 4,157	434 4,157	0.1%	0.6%	0.1%	0.1%	0.3%	4.3%		
389	389	0.1%	18.7%	0.1%	0.1%	0.3%	0.1%		
764	770	0.2%	12.1%	0.4%	0.1%	2.9%	2.0%		
594	588	0.1%	0.3%	0.1%	0.2%	0.1%	0.1%		
		0.1%	98.4%	0.0%	0.0%	0.1%	0.0%		
		0.1%	0.2%	0.0%	0.1%	0.1%	0.0%	0.0%	
	No	ot real data. 20 ata	used for illustrativ	e purposes on	ıly. 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

Prescriptive Analytics - Scenarios

• Plant 1 – MAJOR plant with high volume

1400000

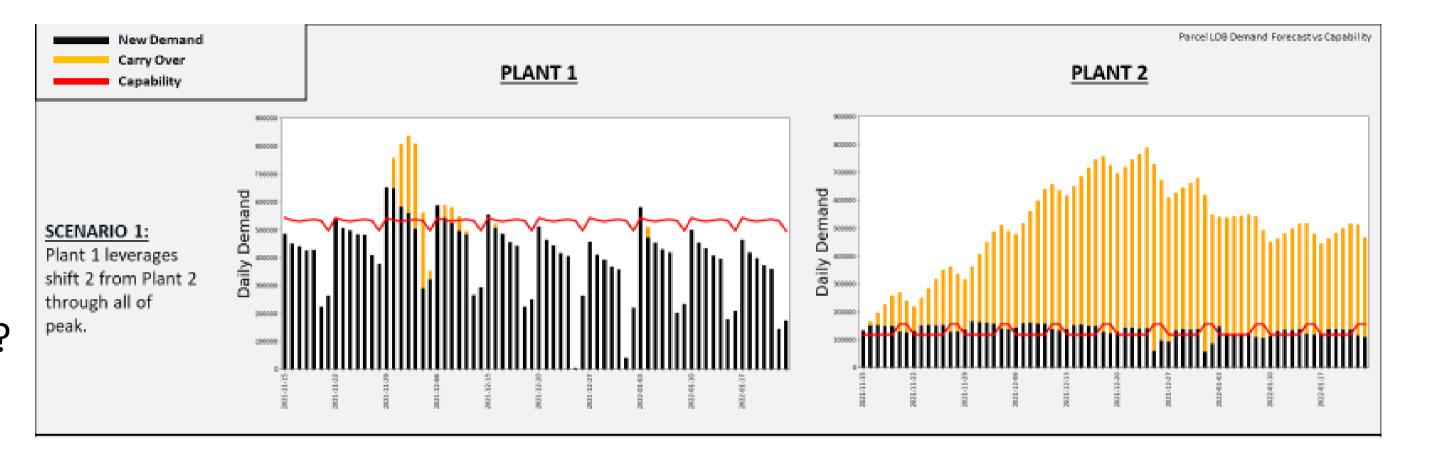
- # of backlog days
- When will backlog clear?





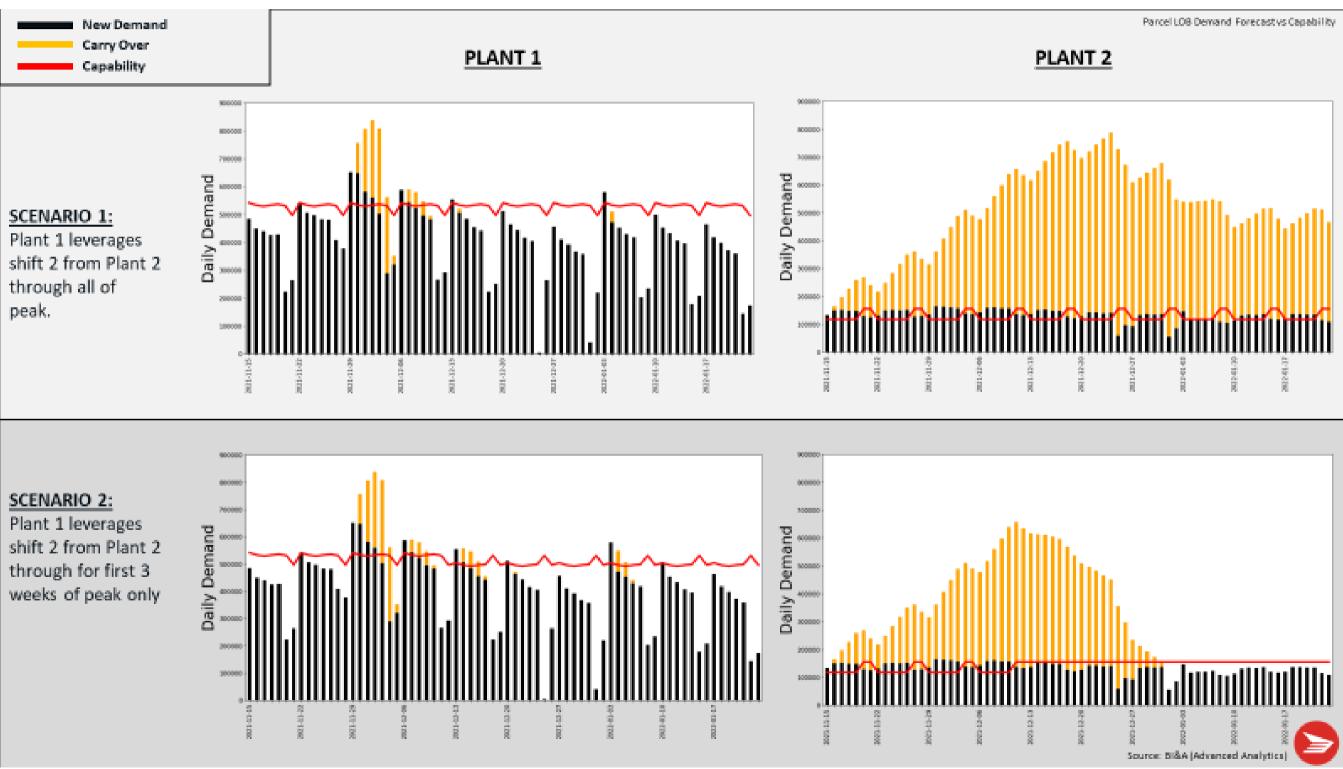
Prescriptive Analytics - Scenarios

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

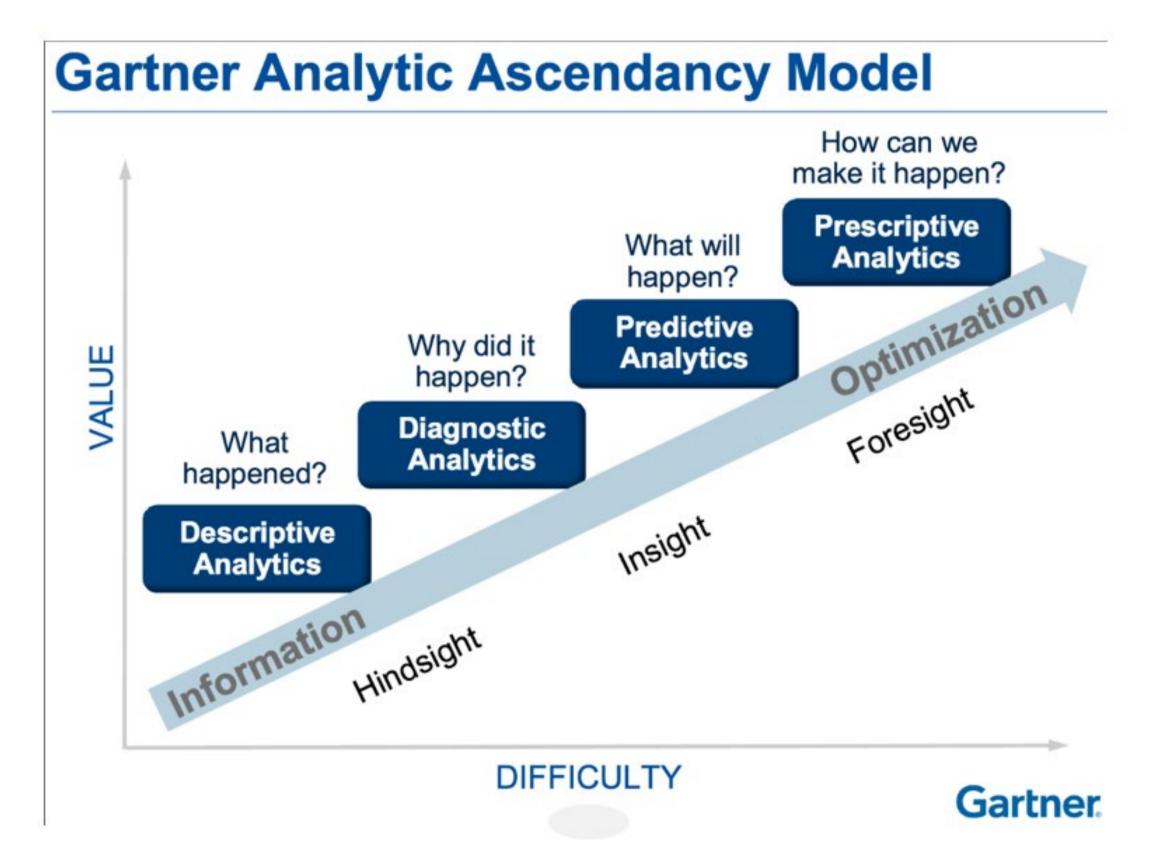


Prescriptive Analytics - Scenarios

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



What's Beyond Prescriptive?







Integrated Analytics

- 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

POSTES
POST
CANADA
CANADA

Carol Wilson,
Director of Advanced Analytics





September 21st, 2023

5pm - 9pm

Ottawa | Lansdowne Horticulture Building

IN PARTNERSHIP WITH