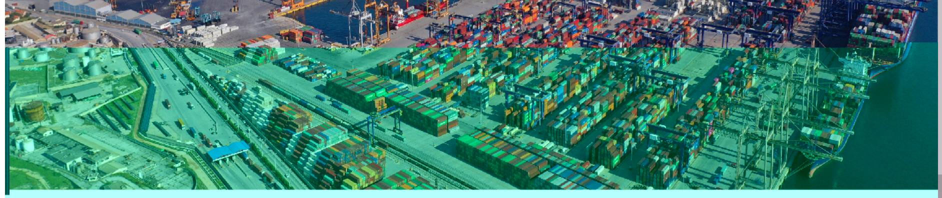


IPIC 2021 | 8th International Physical Internet Conference

Georgia Tech Supply Chain and Logistics Institute
Interdisciplinary Research Center

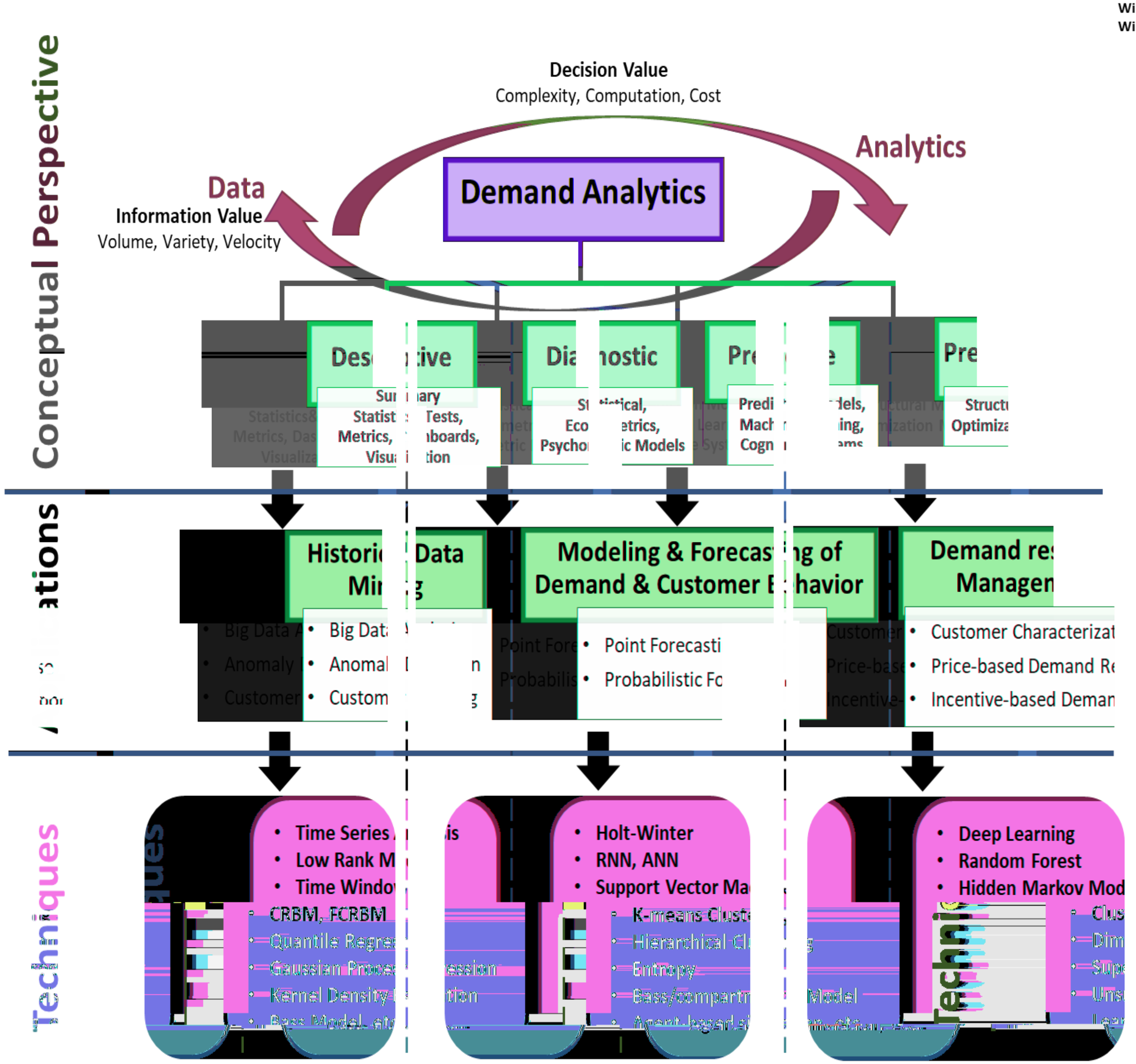
Physical Internet Center
Supply Chain & Logistics Institute

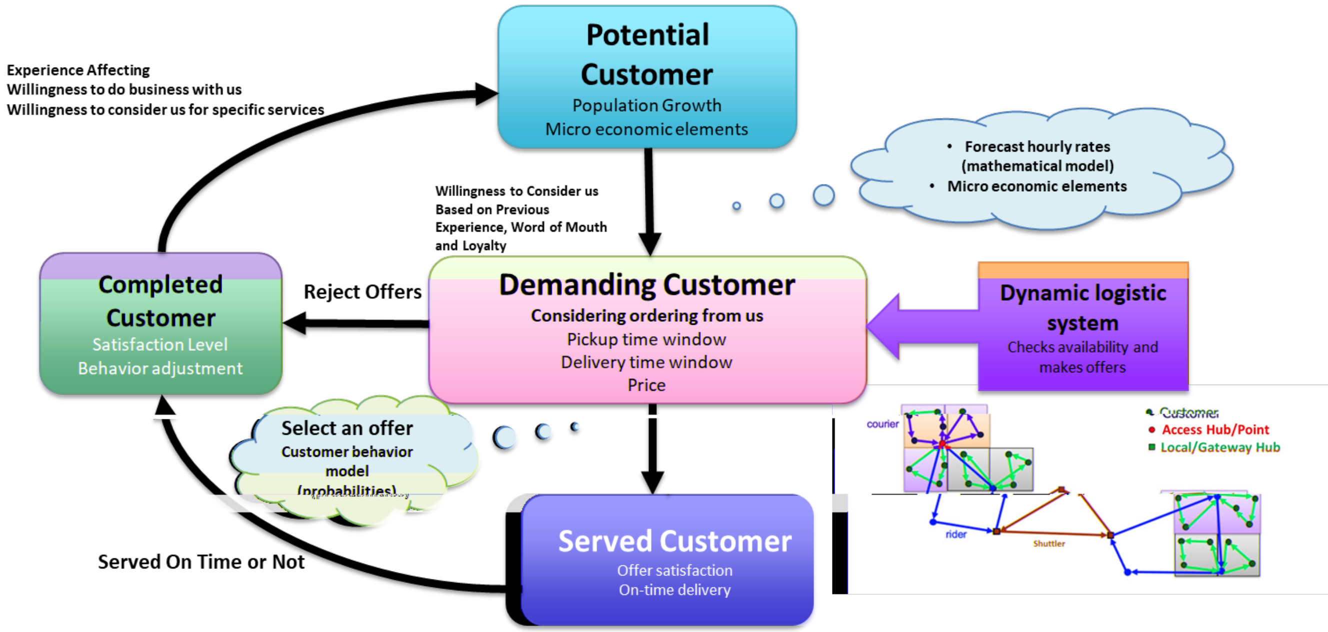
Machine Learning

Predictive Demand Modeling for New Services in Hyperconnected Urban Parcel Logistics

Author & Speaker: Zeynab Bahrami-Bidoni
Graduate Research Assistant and PhD Student
Co-Author: Prof. Benoit Montreuil
Physical Internet Center, ISyE, Georgia Institute of Technology

Literature review: Conceptual Perspective to Applications

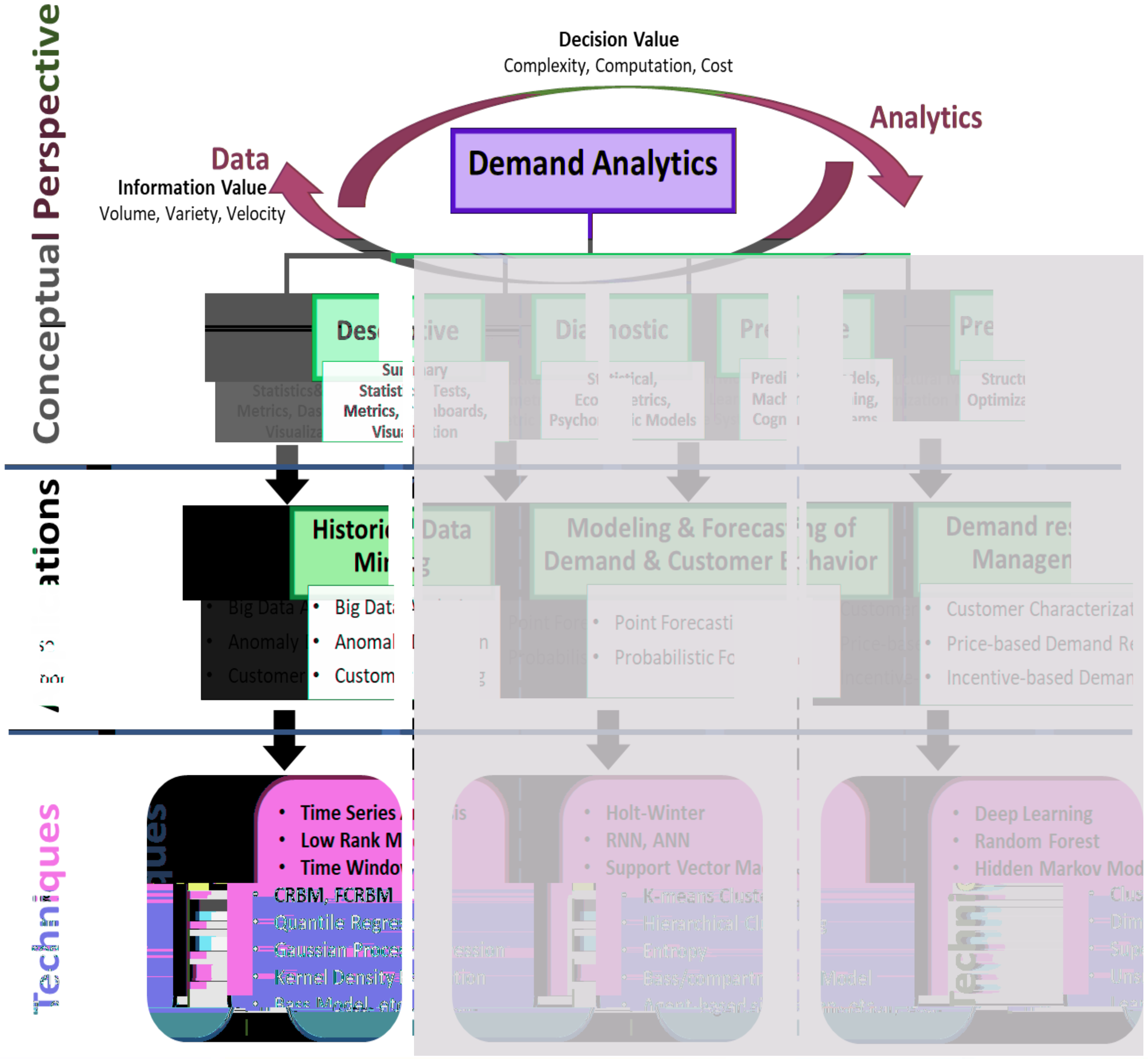




Analytics is known as the scientific process of transforming data into insights for making better decisions. It is commonly dissected into four stages:

- **Descriptive** data summarization and visualization for exploratory purposes,
- **Diagnostic** explanatory models that estimate relationships between variables and allow for hypothesis testing,
- **Predictive** models that enable forecasts of variables of interest and simulation of the effect of marketing control settings,
- **Prescriptive** optimization models that are used to determine optimal levels of marketing control variables.

Descriptive Analytics: Historical Data Mining



A. Big Data Analysis:

- Volume (from terabytes to petabytes)
- Velocity (from one-time snapshots to high-frequency and streaming data)
- Variety (numeric, Long/Lat info, Waybills, and Barcode Scanning streaming)
- Veracity (reliability and validity).

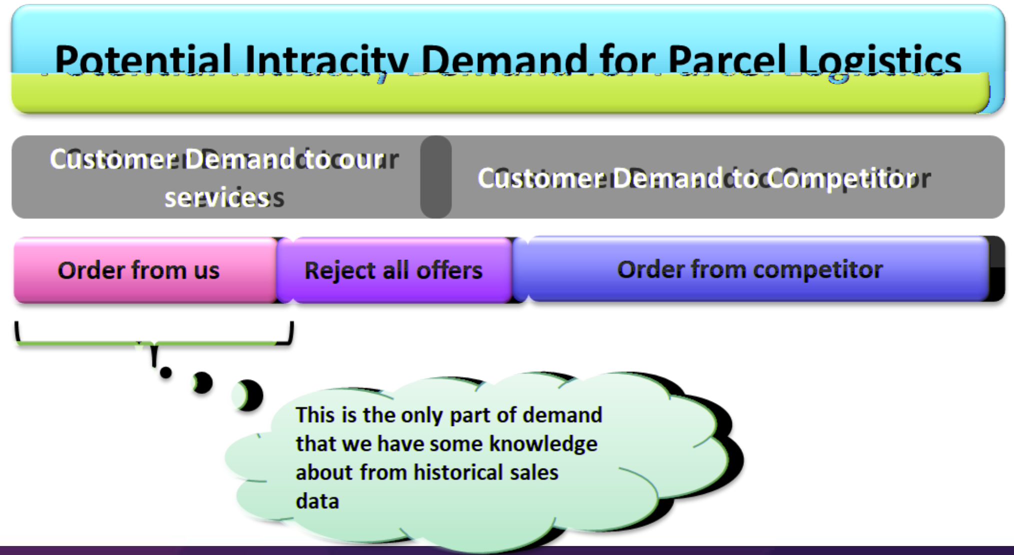
B. Bad Data and Anomaly Detection

- Cleaning dirty data – recognizing outliers/anomalies
- Estimating Missed/Null data

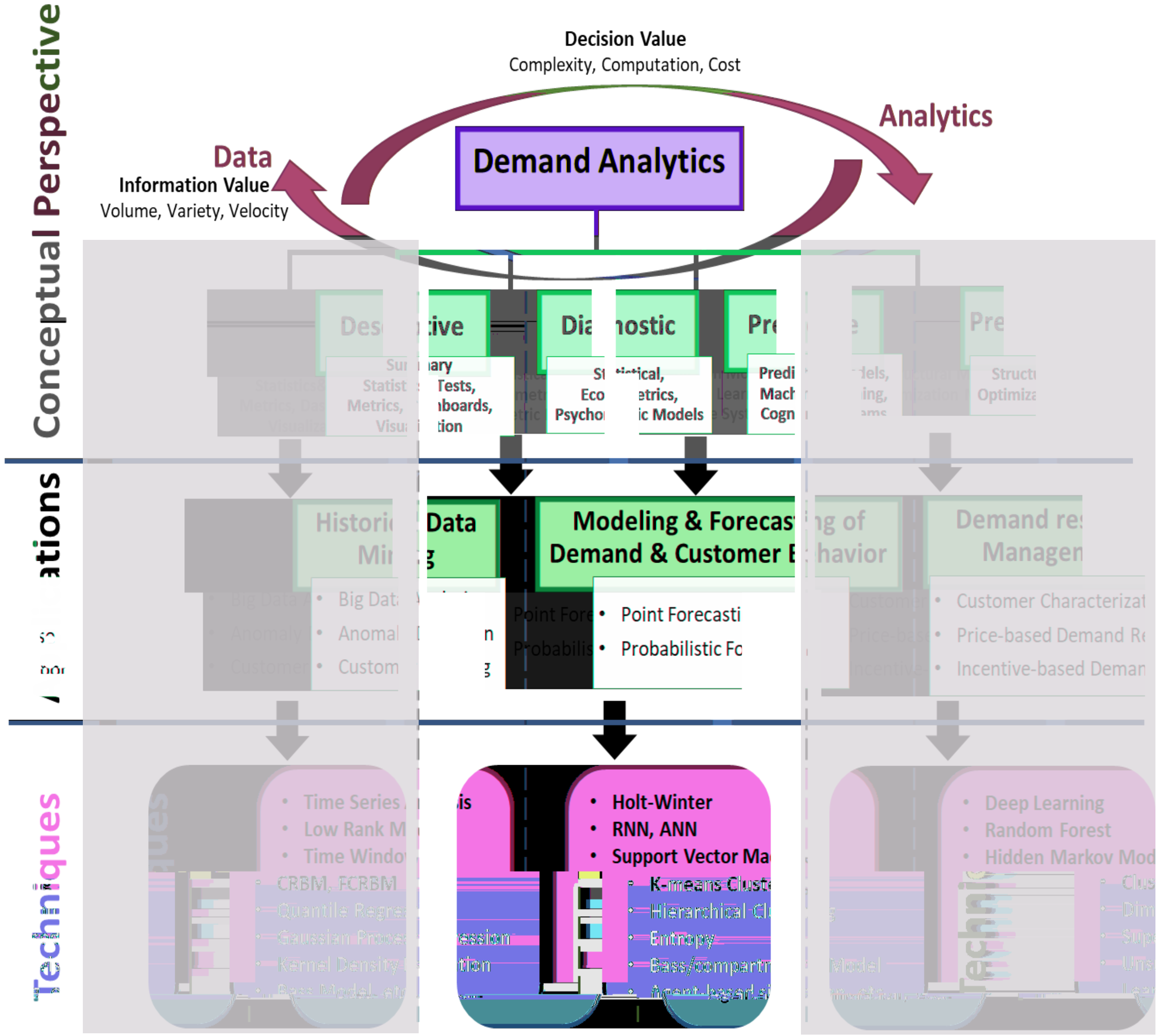
C. Customer profiling

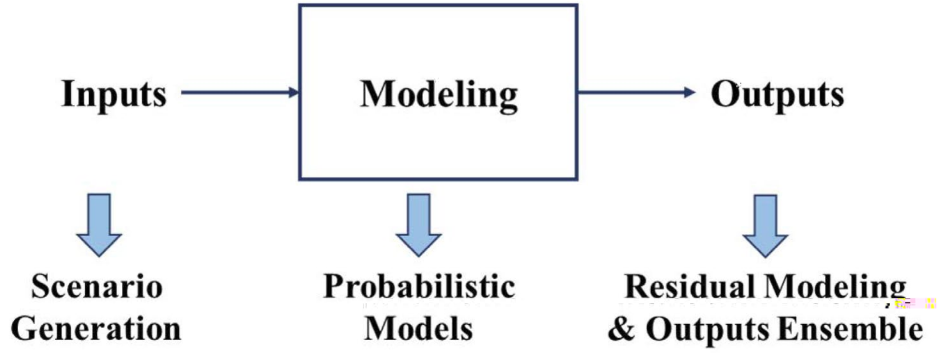
- Direct Clustering
- Indirect Clustering

D. Lack of information Challenges



Diagnostic & Predictive Analytics: Modeling & Forecasting of Demand and Customer Behavior





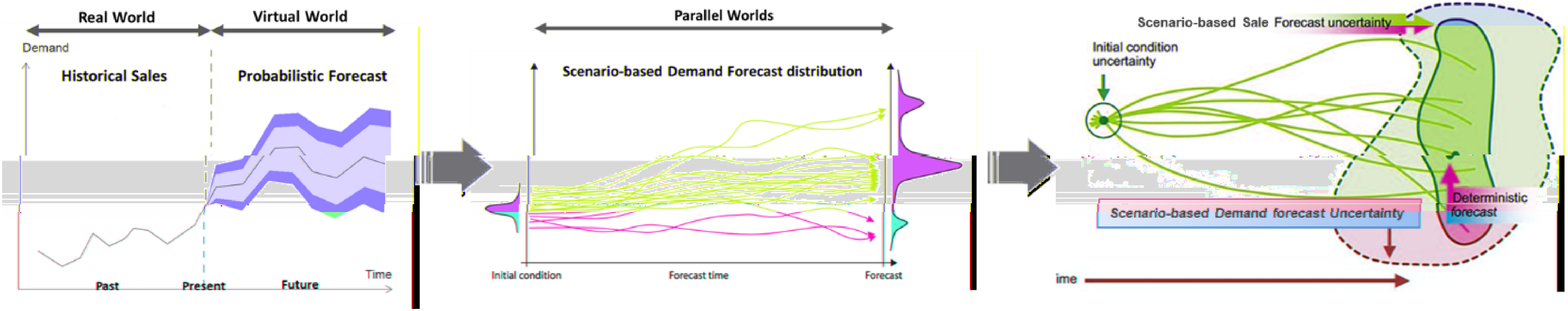
Point Forecasting Vs Probabilistic Forecasting

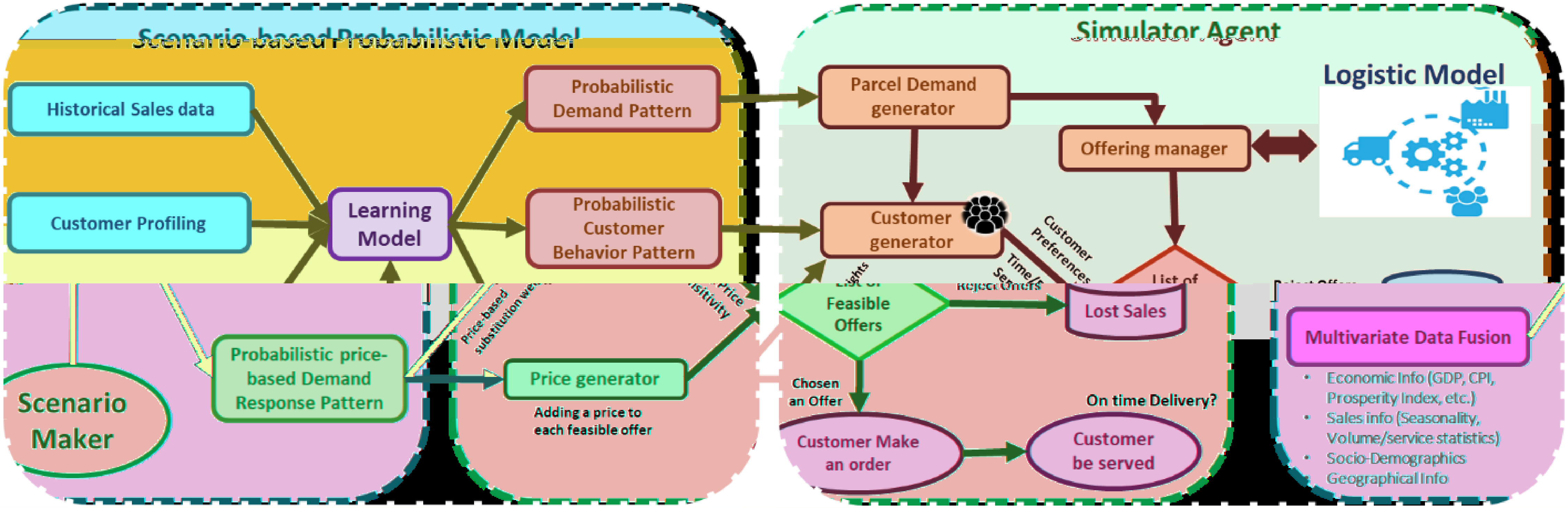
Three ways to modify the workflow to generate probabilistic forecasts:

1) generating multiple input scenarios to feed to a point forecasting model;
2) applying probabilistic forecasting models, such as Gaussian Process regression and kernel Density estimation ;
3) augmenting point outputs to probabilistic outputs by imposing simulated or modeled residuals or making ensembles of point forecasts such as quantile regression .

We have used combination of first two ways for generating scenario-based demand forecast distribution

General Perspective of Proposed Model for Scenario-based Demand and Sale Forecasting Uncertainty





Probabilistic pattern for predicting hourly total demand volume

F(y, m, w, d, wd, h)= P(y) * G(w) * D(m,d) * T(h,wd)

P(y)= Total demand volume in year y considering growth factors of potential customer population
G(w)= Weekly residual load variation
D(m,d) = day-month template structure (daily portion pattern of week load on different dates)
T(h,wd) = hour-weekday template structure (hourly portion pattern of daily load on different week days)

y=year 2017,2018,…, 2035
m=month 1,2,…,12
w=week 1,2,…52
d=date 1,2,…,31
wd=weekday Sunday (1), Monday(2),…,Saturday(7)
h=hour 1,2,…,23

Function F(y, m, w, d, wd, h) is a probabilistic function which estimate demand volume in hour h of day d in month m of year y (within week w of year & weekday wd)

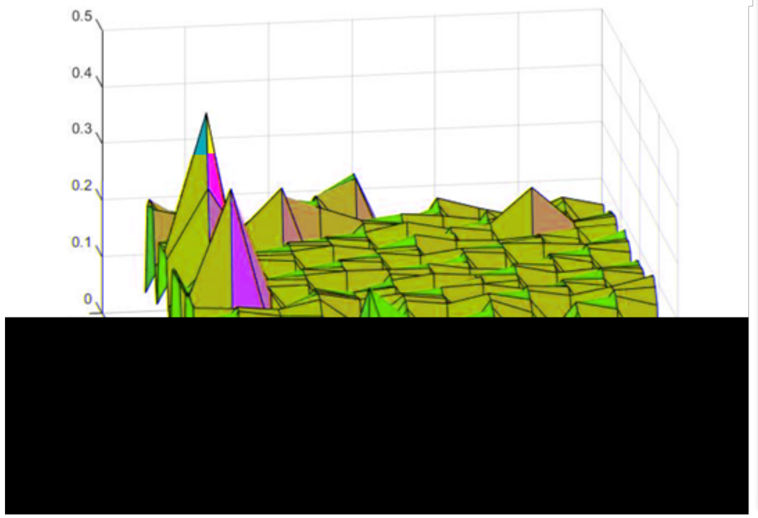
$$
\begin{aligned}
T(h,d) = {} & B_1 \cos\left(\frac{\pi h}{24} + B_2\right)\cos\left(\frac{\pi d}{7} + B_3\right) + B_4 \cos\left(\frac{2\pi h}{24} + B_5\right)\cos\left(\frac{\pi d}{7} + B_6\right) + B_7 \cos\left(\frac{3\pi h}{24} + B_8\right)\cos\left(\frac{\pi d}{7} + B_9\right) + \ldots + B_{38} \cos\left(\frac{\pi h}{24} + B_{39}\right)\cos\left(\frac{9\pi d}{7} + B_{40}\right) + \ldots \\
& + B_{44} \sin\left(\frac{\pi h}{24} + B_{45}\right)\sin\left(\frac{\pi d}{7} + B_{46}\right) + B_{47} \sin\left(\frac{2\pi h}{24} + B_{48}\right)\sin\left(\frac{\pi d}{7} + B_{49}\right) + B_{50} \sin\left(\frac{3\pi h}{24} + B_{51}\right)\sin\left(\frac{\pi d}{7} + B_{52}\right) + \ldots + B_{83} \sin\left(\frac{\pi h}{24} + B_{84}\right)\sin\left(\frac{9\pi d}{7} + B_{85}\right) + \ldots \\
& + B_{91} \sin(B_{92} * h + B_{93})\cos(B_{94} * d + B_{95}) + B_{96} \cos(B_{97} * h + B_{98})\sin(B_{99} * d + B_{100}) + B_{101} \sin(B_{102} * h + B_{103}) + B_{104} \cos(B_{105} * h + B_{106}) + \\
& + B_{107} \sin(B_{108} * d + B_{109}) + B_{44} \cos(B_{110} * d + B_{111}) + B(112)
\end{aligned}
$$

$$h = 1, \ldots, 24$$

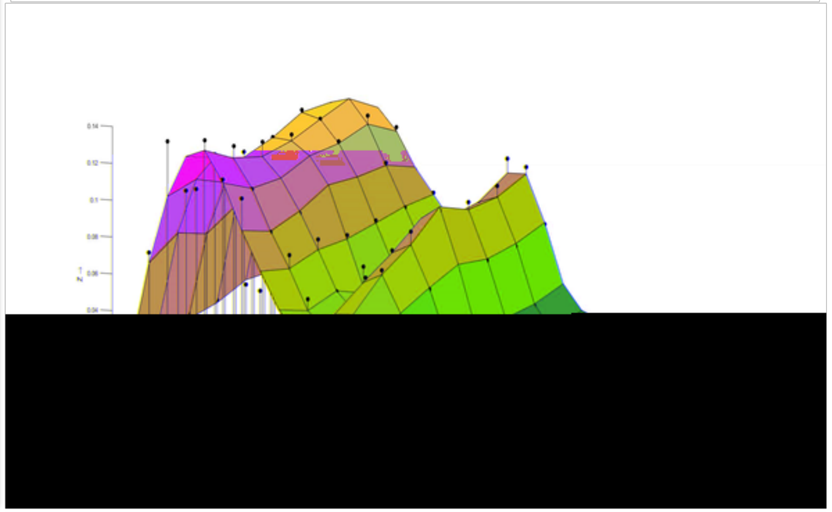
$$d = 1, \ldots, 7 \qquad 1 = Sunday \qquad 7 = Saturday$$

- **In order to model weekdays-hourly template structure we do Surface Fitting on historical dataset and compute coefficients with %95 confidence**

(a) Two-Dimensional Days-Months variation surface fractional rates (daily week load portion pattern)



(b) Two-Dimensional Weekdays-Hourly variation surface fractional rates (daily week load portion pattern)



Scenario-based demand estimation for new faster services

Input pattern

Based on daily average rates from historical data

Category • Origin/Destination Distance • Source Type • Destination Type	Time factor (hour of day) 1:00AM to 24:00PM	% Sales for Service #1 Delivery within 10 Hours	% Sales for Service #2 Delivery within 18 Hours	% Sales for Service #3 Delivery within 30 Hours	% Sales for Service #4 Delivery within 70 Hours

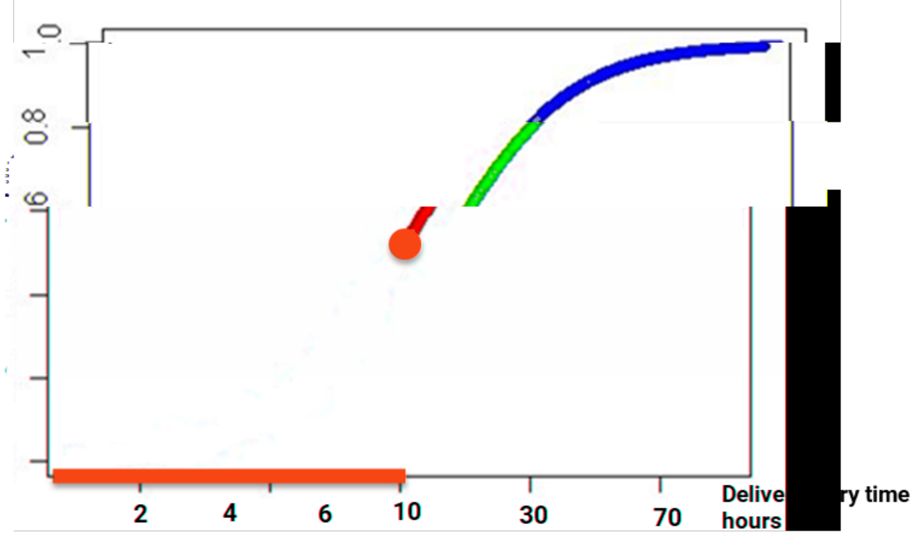
New faster Services

Service #5: Delivery less than an Hour
Service #6: Delivery within 2 Hours
Service #7: Delivery within 4 Hours
Service #8: Delivery within 6 Hours

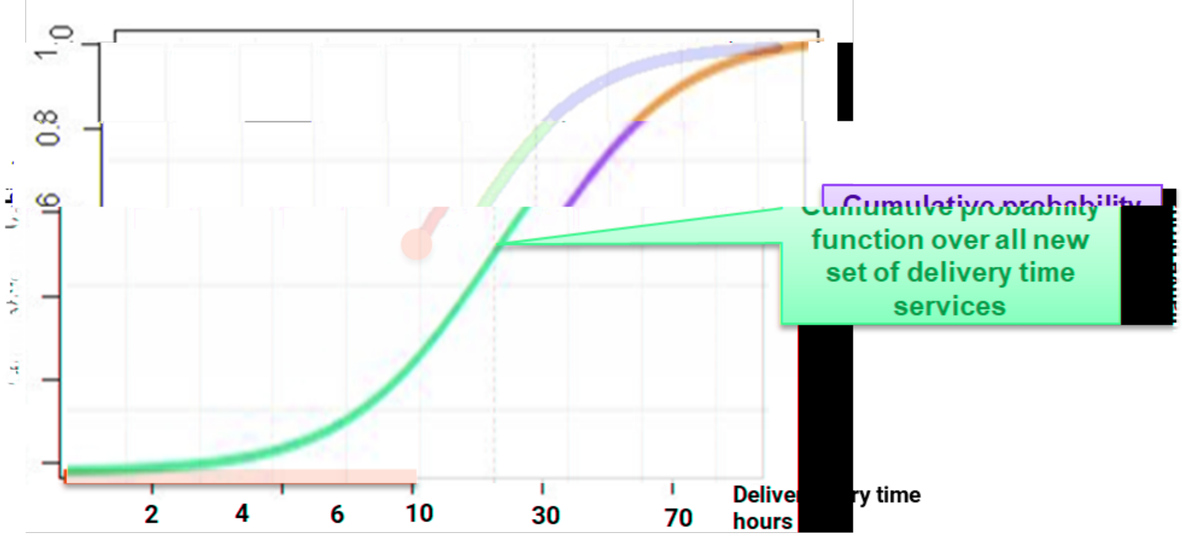
TS: Delivery-time sensitivity weight

e.g. TS=(sales rate on service#1) / (sales on rate service#1+ sales rate on service#2)

Cumulative probability function of demand in same order category extracted from historical sales data over old set of delivery-time service offers



Transformed cumulative probability function of demand in same order category resulted from Algorithm # 1 over new set of delivery-time services including new faster offers



Algorithm for simulation of customer preference over all potential continuous delivery-time

D: *Average daily demand forecasted by first model*

P: *Total percentage of potential customers seeking new faster services (one of scenario assumption).*

$O^{Old} = \{t_1^{Old}, t_2^{Old}, \ldots, t_s^{Old}\}$ *: The set of all promised delivery-times in the old service offering system.*

$O^{New} = \{t_1^{New}, t_2^{New}, \ldots, t_r^{New}\}$ *: The set of all faster delivery-times in the new service offering system.*

$V^c = \left(v^c_{t_1^{Old}}, v^c_{t_2^{Old}}, \ldots, v^c_{t_s^{Old}}\right)$, $\sum_{i=1}^{s} v^c_{t_i^{Old}} = 1$ *: Probability vector of demand over the set of old offering services.*

Bass Model (1969) was defined using a differential equation in which purchases are initiated by mass communication and further driven by word of mouth (WOM) from past purchases.

Algorithm #1:

Goal: *Computing continues cumulative probability function of category c for selecting over all delivery-time services.*

Stage 1: Compute the vectors $K^c = \left(K_0^c, K_1^c, K_2^c, \ldots, K_s^c\right)$,

where $K_0^c = P * D * w_{pd}^c$ *and*

$K_i = (1-P) * D * w_{pd}^c * v^c_{t_i^{Old}}$, $i = 1, \ldots, s$

Stage 2: *Normalize* K^c *to get the vector* $K^{c'} = \dfrac{1}{\sum_{i=0}^{s} K_i^c} K^c$

where its first component $K_1^{c'}$ *is the cumulative probability of demand over all new offering services*

Stage 3: $a = K_1^{c'} / \left(t_r^{New}\right)^2$, $L = K_1^{c'}$

Do for $i = 1:r$
$\hat{F}(t_i^{New}) = a * (t_i^{New})^2$
End

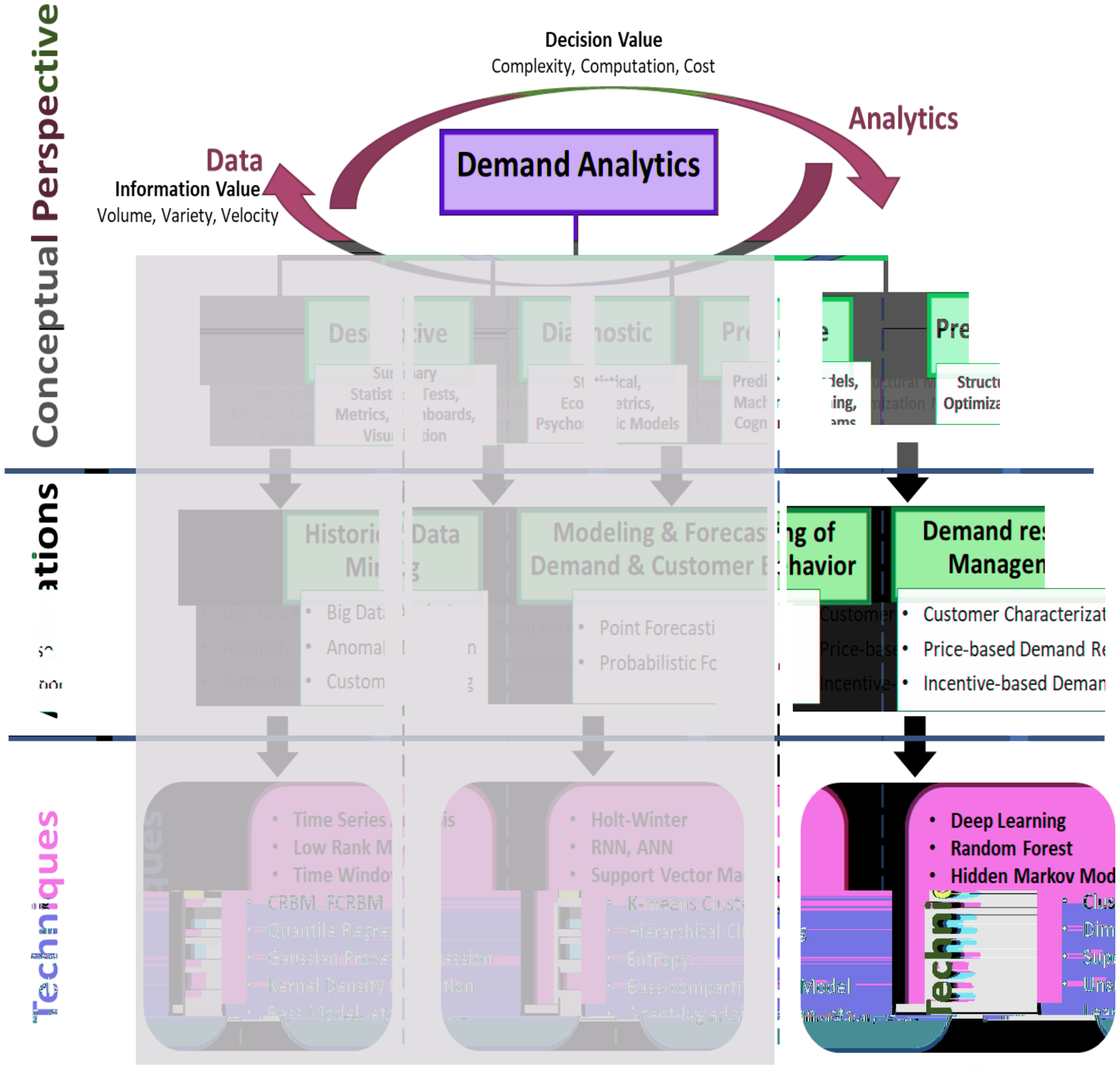
and

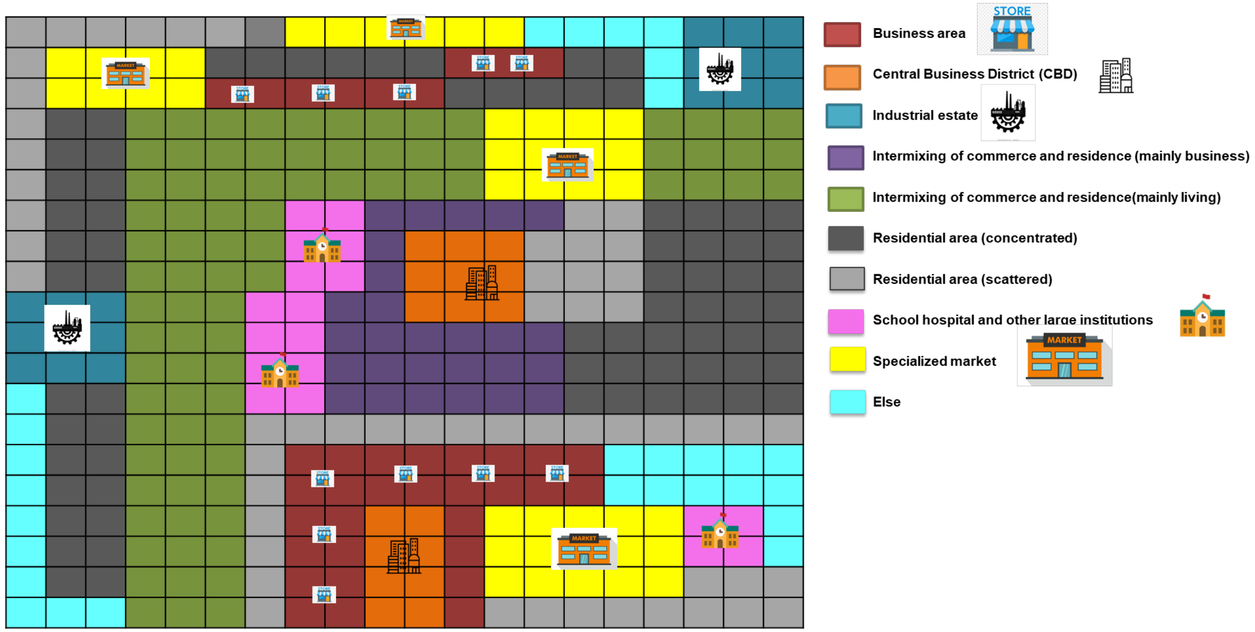
Do for $i = 1:s$
$L = L + K_{i+r}^{c'}$
$\hat{F}(t_i^{Old}) = L$
End

Stage 4: *Fitting below function inspired by Bass diffusion Model*
$F_{p,q}(t) = \left(1 - \exp(-(p+q)t)\right) / \left(1 + (q/p)\exp(-(p+q)t)\right)$ *and compute the optimal* p^*, q^* *coefficients with lower function error to the data sets* $T = \{t,\ t \in (O^{New} \cup O^{Old})\}$, $\hat{F}(T) = \{\hat{F}(t),\ t \in T\}$

[redacted]

Prescriptive Analytics: Demand Response Management





A. Customer Characterization:

- Geographical attributes
- Economic attributes (GDP, CPI, Prosperity Index, etc.)
- Socio-demographics

B. Price-based demand response analysis

- Price-based Sensitivities of Customer preferences
- Optimal dynamic pricing

C. Incentive-based demand response analysis

- Customer Satisfaction level
- WOM (word of Mouth)
- Advertisements-incentive offers/plans

Prescriptive Analytics: Price/Incentive Demand Response Analytics

F(t) and *f(t)* are the cumulative and non-cumulative proportions of demand at offered delivery-time t,
Y(t) is the total number of potential customers demanding faster services up to but not including offers with delivery time t.
(p) is the rate of spontaneous demand adoption and (q) is the rate of imitation of demand adoption (the optimal p and q parameters for any order category computed by the algorithm #1).
x(t), P(t) , and A(t) are respectively the market effort, the price and the advertising for the ith service offer with t promised delivery-time. These variables enter the market effort equation as percentage increases.

$$\frac{f(t_i)}{1-F(t_i)} = (p+qY(t_i))x(t_i), \quad x(t_i) = 1+\alpha_1 \frac{P(t_i)-P(t_{i-1})}{P(t_{i-1})} + \alpha_2 \frac{\max\{0, A(t_i)-A(t_{i-1})\}}{A(t_{i-1})}, \quad t_i \in T$$

$$S(t_i) = F(t_i) - F(t_{i-1}) + e$$

Where S(t) is the sale on offered service with delivery-time t and e is an additive normally distributed error term and F(t) is given by equation:

$$F(t_i) = \frac{1-\exp\{-\bar{X}(t_i)(p+q)\}}{1+(q/p)\exp\{-\bar{X}(t_i)(p+q)\}}, \qquad \bar{X}(t_i) = t + \alpha_1 Ln\left(\frac{P(t_i)}{P^{\min}}\right) + \alpha_2 Ln\left(\frac{\tilde{A}(t_i)}{A(0)}\right), \qquad t_i \in T$$

$$t^* = (\ln q - \ln p)/(p+q)$$

Scenario-based User-Interactive Demand Generator APP

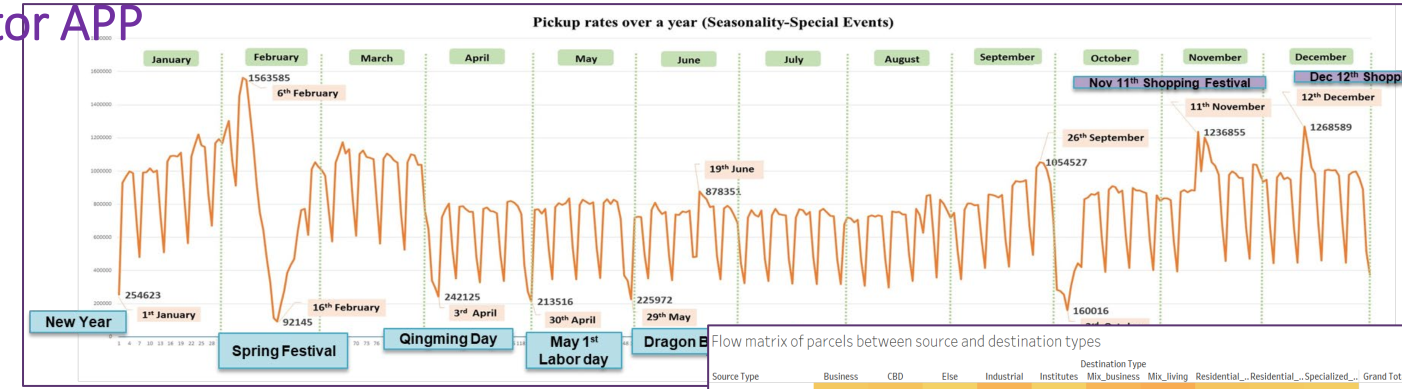
➢ Inputs based on historical patterns

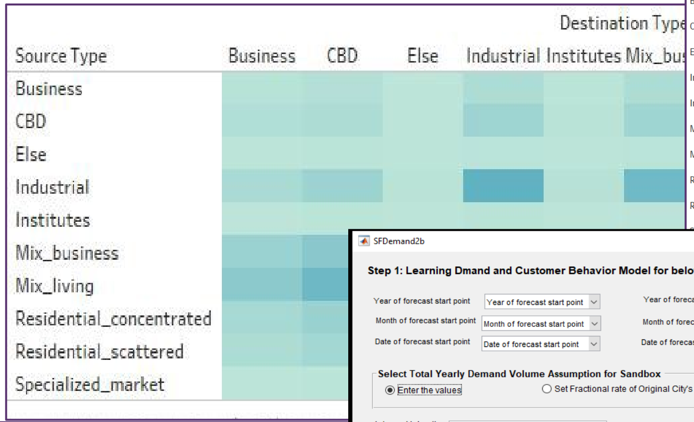
By setting some assumption in App interface will make a scenario. Input patterns by csv files will be given to the App are including:

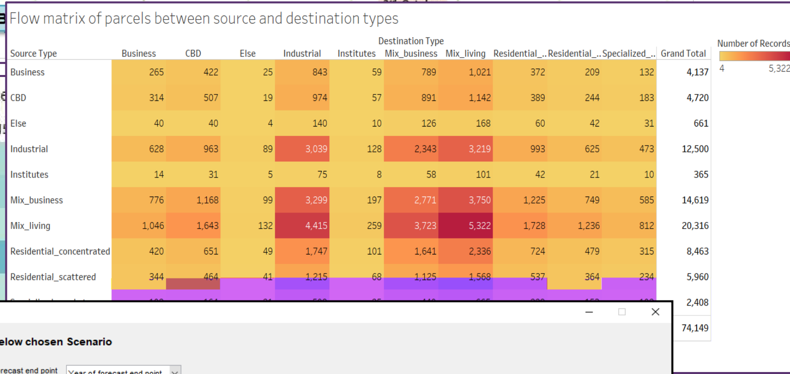
- Seasonality variations and special events
- Customer preferences on old set of available service levels
- Fractional Distribution of demand volume over all origin/destination pairs
- Geo-characterizations info and rate of their demand on different time factors

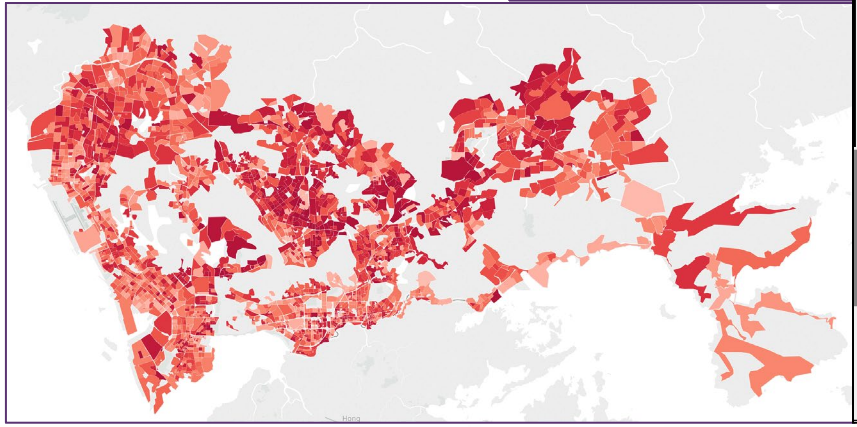
➢ Outputs of App

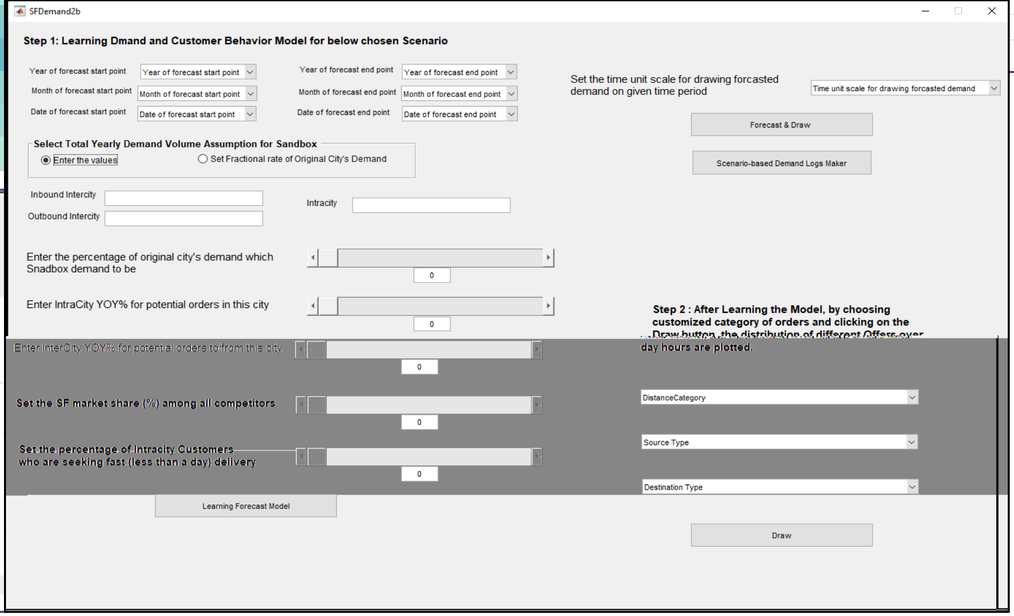
- scenario-based probabilistic models for predicting demand volume and customer behavior to be transposed into orders in a simulated logistics system.
- scenario-based forecasted demand logs for an arbitrary time horizon











Outputs' format of scenario-based demand Logs maker tool

- **Intracity demand logs**

Year	Month	Day	Hour	Minute	Second	SourceZone	DestinationZone	Service	DistanceCategory	Source_Type	Destination_Type
2020	4	29	0	1	1	ZoneCN011	ZoneMML001	NextDay	category2	Business	Mix_living
2020	4	29	0	1	9	ZoneD030	ZoneMUL001	NextDay	category2	Residential_concentrated	Mix_living
2020	4	29	0	1	12	ZoneFQ013	ZoneAD007	NextDay	category1	Mix_business	Specialized_market
2020	4	29	0	1	17	ZoneJ010	ZoneDF007	NextDay	category2	Mix_business	Mix_living
2020	4	29	0	1	18	ZoneA002	ZoneDJ005	sixHours	category2	Business	Industrial
2020	4	29	0	1	23	ZoneAJ011	ZoneM013	NextDay	category2	Industrial	Mix_living
2020	4	29	0	1	23	ZoneUAL01	ZoneAT029	fourHours	category2	Mix_living	CBD
2020	4	29	0	1	23	ZoneG011	ZoneU013	NextDay	category2	CBD	CBD
2020	4	29	0	1	32	ZoneBN013	ZoneCM007	NextDay	category2	Mix_business	Mix_living

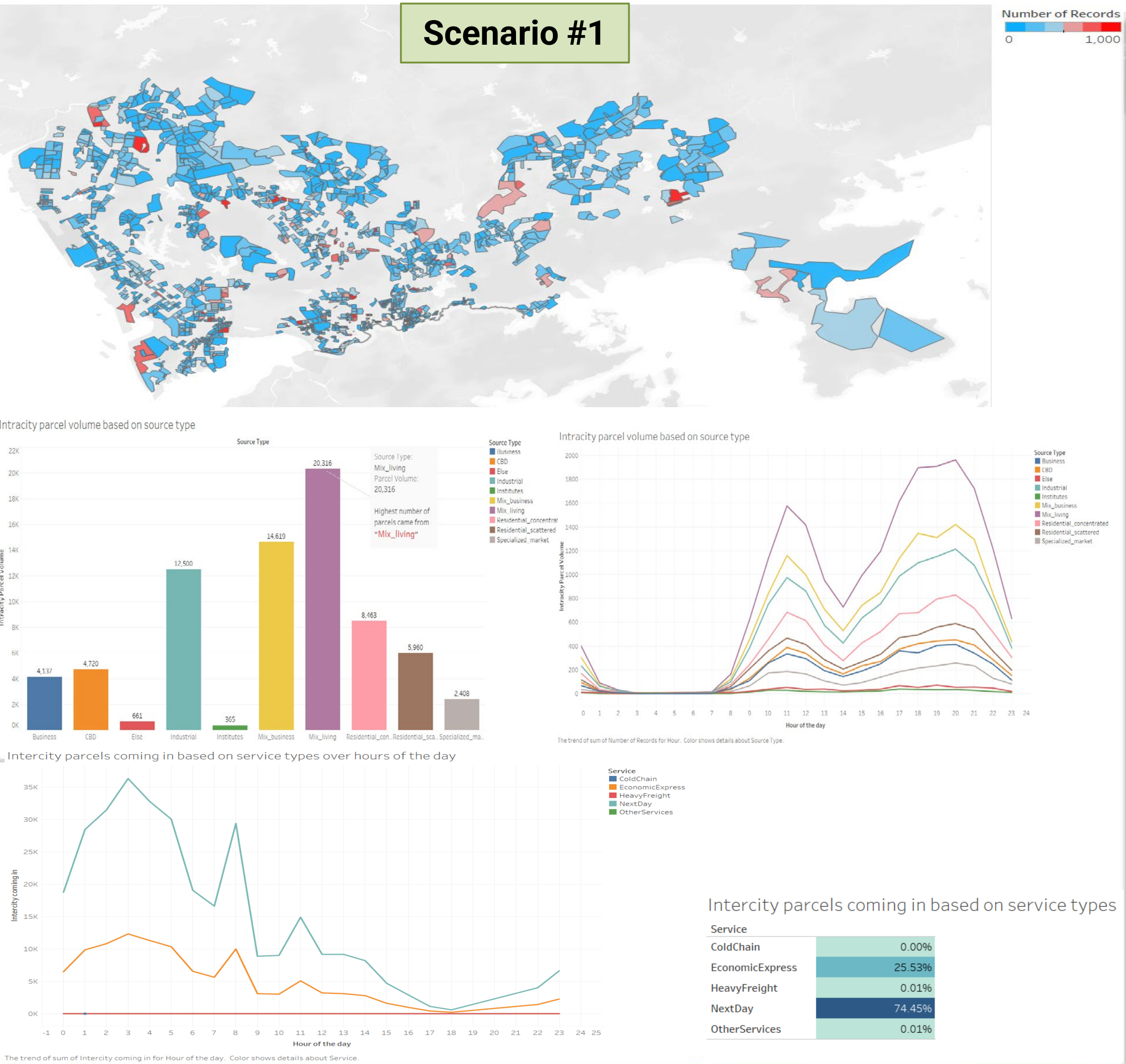
- **Inbound Intercity demand logs**

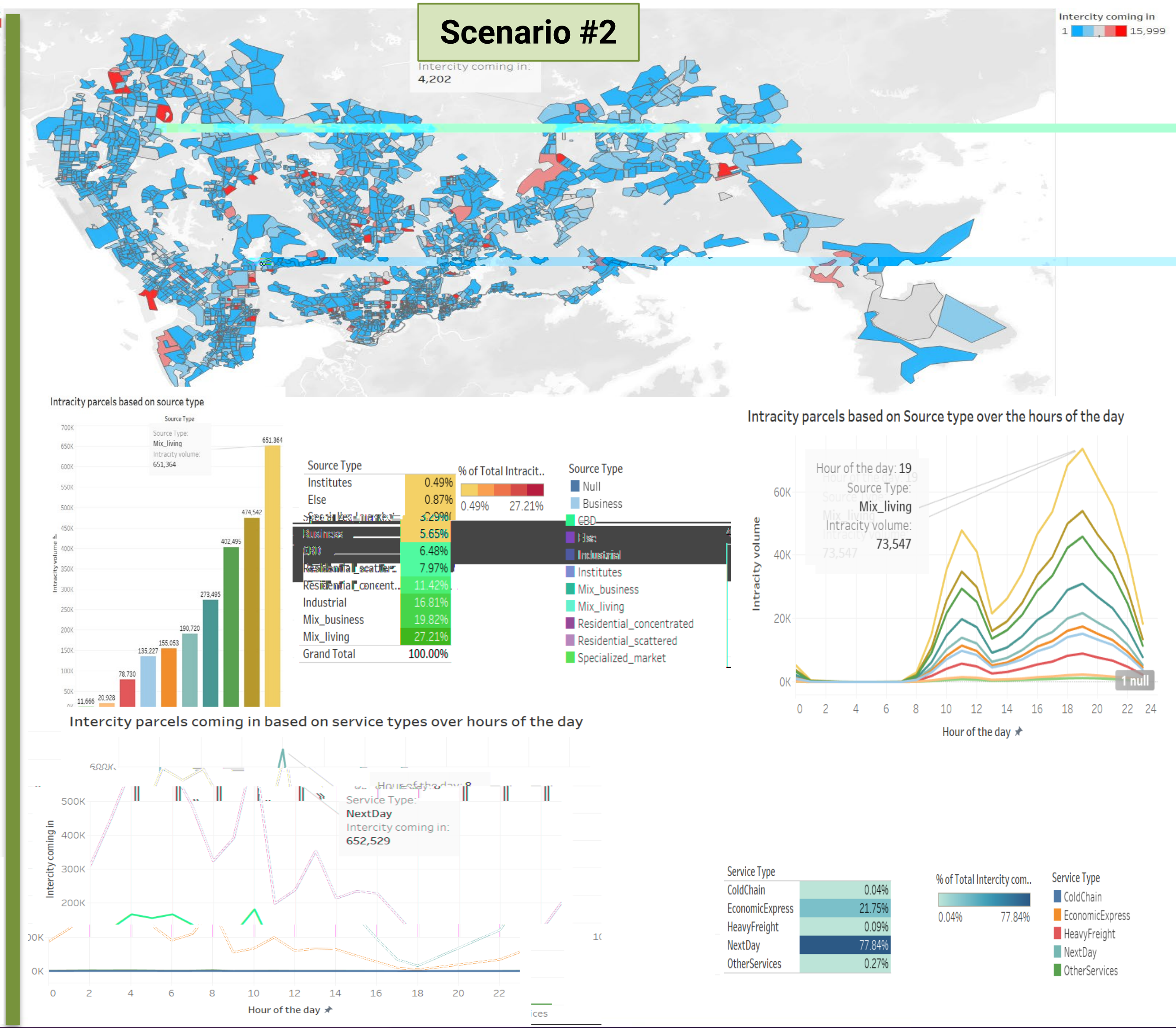
Year	Month	Day	Hour	Minute	second	SourceCity_Code	Delivery_GatewayHub	DestinationZone	Service
2020	4	29	0	1	1	762	GateHub#1	ZoneMOL002	EconomicExpress
2020	4	29	0	1	1	512	GateHub#2	ZoneJ021	NextDay
2020	4	29	0	1	1	769	GateHub#1	ZoneEA005	EconomicExpress
2020	4	29	0	1	1	28	GateHub#1	ZoneEP012	NextDay
2020	4	29	0	1	1	20	GateHub#2	ZoneBH025	NextDay
2020	4	29	0	1	2	20	GateHub#2	ZoneAN006	NextDay
2020	4	29	0	1	2	595	GateHub#3	ZoneCP012	NextDay
2020	4	29	0	1	2	415	GateHub#3	ZoneBK009	NextDay

- **Outbound Intercity demand logs**

Year	Month	Day	Hour	Minute	second	SourceZone	Pickup_GatewayHub	DestinationCity_Code	Service
2020	4	29	0	1	1	ZoneFP013	GateHub#1	571	NextDay
2020	4	29	0	1	1	ZoneN019	GateHub#2	730	EconomicExpress
2020	4	29	0	1	1	ZoneFK004	GateHub#2	913	NextDay
2020	4	29	0	1	1	ZoneCP008	GateHub#3	531	NextDay
2020	4	29	0	1	2	ZoneAE026	GateHub#2	553	EconomicExpress
2020	4	29	0	1	2	ZoneAL012	GateHub#2	753	NextDay
2020	4	29	0	1	2	ZoneFJ033	GateHub#2	838	EconomicExpress
2020	4	29	0	1	3	ZoneAG048	GateHub#3	769	EconomicExpress
2020	4	29	0	1	3	ZoneT032	GateHub#3	25	EconomicExpress

Comparison between demand Logs corresponding two example scenarios





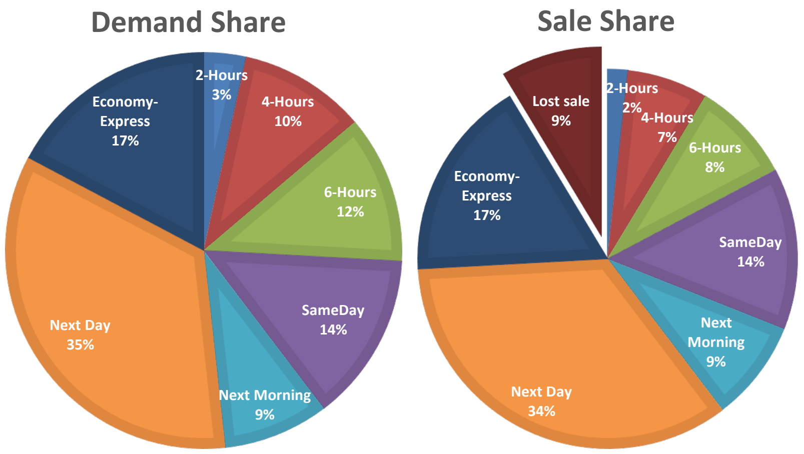
Analysis over simulator results by feeding scenario-based probabilistic patterns for demand and customer behavior

Type 1 comparison:

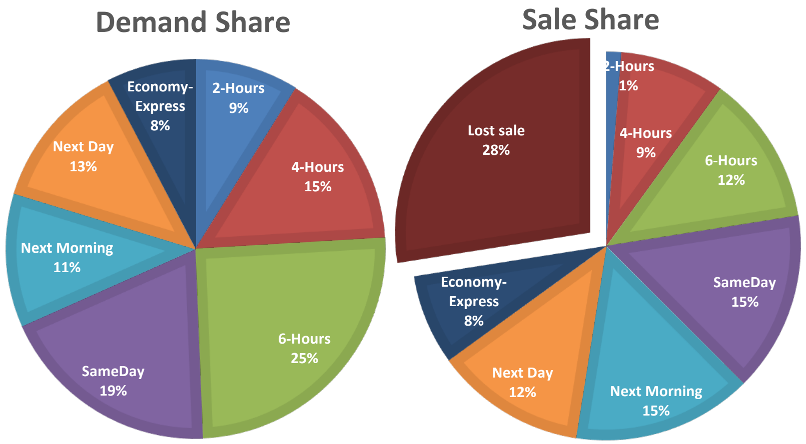
Comparing the simulated performance of a designed logistic system on multi-scenario demand.

For current logistic design in this example, the level of customer satisfaction on demand scenario #2, is lower and lost sale is higher. We can consider some decision policy for compensating the lack of resources to avoid lost sales in scenario #2.

Scenario #1



Scenario #2

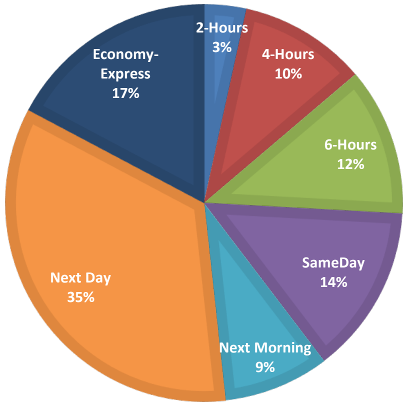


Type 2 comparison :

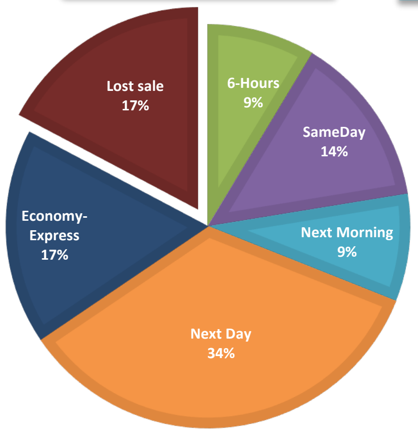
Comparing the performance of different logistic system configurations on a demand scenario.

In this example shows that logistic configuration of new hyperconnected multi-tier mesh network has a better performance and reached higher Level of customer satisfaction and lower lost sales compare to other logistic designs.

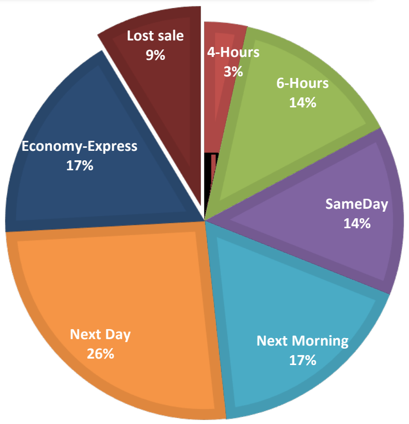
Demand Scenario



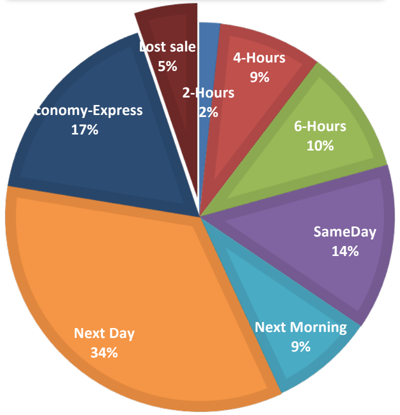
Hub-and-Spoke Network Design



Current Hyperconnected Multi-tier Mesh Network



New Hyperconnected Multi-tier Mesh Network



Contributions and Conclusion

Benefits/Outputs of proposed Scenario-based App for demand and customer behavior modeling & forecasting

- Ability to make a wide rage of various scenarios
- Ability to simulate/forecast customer preference on new services
- Resulting scenario-based probabilistic patterns which are feedable to the logistic simulation for testing performance and sales attributes including:
  - Three Scenario-based probabilistic patterns for hourly total demand volume in future: Intracity pattern, Inbound Intercity pattern, Outbound Intercity pattern
  - Probability distribution of customer preferences on different offers for all categories
- Providing scenario-based forecasted demand logs for any arbitrary time horizon.

Benefits of Comparison analysis over scenarios:

- ❖ Comparison multi-scenarios by analysis on demand logs

  Providing deep managerial insights on demand shape and its geographical distribution in terms of volume and service types which is helpful to be prepared for future risks/challenges with suitable policies

- ❖ Comparison multi-scenarios by analysis on simulated sales after feeding probabilistic patterns leads to
  - Better understanding of sale shape and its geographical distribution in terms of volume and service types
  - Better understanding about required capacities or lack of resources in different locations or different hours of day which result to lost sales
  - Comparing logistic performance for different Hub designs
  - Comparing the optimality of different routing design
  - Comparing the level of on time delivery and customer satisfactions

Further research: a potential for model improvement is to explicitly account for customer satisfaction and its positive/negative impacts on future demand.


Thank You

References:

- Abedi, V. S. (2019): Compartmental diffusion modeling: Describing customer heterogeneity & communication network to support decisions for new product introductions. Physica A: Statistical Mechanics and its Applications, 536, 120964.
- Bahrami-Bidoni Z., B. Montreuil (2021): Enabling Scientific Assessment of Large Scale Hyperconnected Urban Parcel Logistics: Scenario-based Demand and Customer Behavior Modeling. Proceedings of IISE 2021 Conference.
- Bass, F. M. (1969): A new product growth for model consumer durables. Management science, 15(5), 215-227.
- Campos M., B. Montreuil, L. McGinnis, S. Kaboudvand, S. Kwon, Z. Bahrami-Bidoni, L. Faugere, S. Buckley (2021): Enabling Scientific Assessment of Large Scale Hyperconnected Urban Parcel Logistics: System Configuration and Assessment. Proceedings of IISE 2021.
- Chicco G. (2016): Customer behavior and data analytics. in 2016 International Conference and Exposition on Electrical and Power Engineering (EPE). IEEE, 2016, pp. 771–779.
- Filik, Ü. B., Gerek, Ö. N., Kurban, M. (2011): A novel modeling approach for hourly forecasting of long-term electric energy demand. Energy Conversion and Management, 52(1), 199-211.
- Hariharan, V. G., Talukdar, D., Kwon, C. (2015): Optimal targeting of advertisement for new products with multiple consumer segments. Internat. J. of research in marketing, 32(3), 263-271.
- Hong T., S. Fan (2016): Probabilistic electric load forecasting: A tutorial review. International Journal of Forecasting, vol. 32, no. 3, pp. 914–938, 2016.
- Kaboudvand S., M. Campos, B. Montreuil (2021): Enabling Scientific Assessment of Large Scale Hyperconnected Urban Parcel Logistics: Agent-Based Simulator Design", IISE 2021 Conference.
- Kiesling, E., Günther, M., Stummer, C., Wakolbinger, L. M. (2012): Agent-based simulation of innovation diffusion: a review. Central European Journal of Operations Research, 20(2),183-230.
- Rahmandad H., Sterman J. (2008): Heterogeneity & network structure in the dynamics of diffusion: Comparing agent-based and differential equation models. Management Science, 54(5), 998-1014.
- Rand W., R. T. Rust (2011): Agent-based modeling in marketing: Guidelines for rigor. International Journal of research in Marketing, 28(3), 181-193.
- Ramírez-Hassan, A., Montoya-Blandón, S. (2020). Forecasting from others' experience: Bayesian estimation of the generalized Bass model. International Journal of Forecasting, 36(2), 442-465.
- Wang Y., Q. Chen, T. Hong, C. Kang (2018): Review of smart meter data analytics: Applications, methodologies, and challenges," IEEE Transactions on Smart Grid, v10, no3, 3125–3148.
- Wedel M., P. Kannan (2016): Marketing analytics for data-rich environments," Journal of Marketing, vol. 80, no. 6, pp. 97–121, 2016.
- Yin, P., Dou, G., Lin, X., & Liu, L. (2020): A hybrid method for forecasting new product sales based on fuzzy clustering and deep learning. Kybernetes.