PHARMA MARKETING: THE UPSIDE OF SEGMENTATION USE CASE

DEC 2020

MEET THE TEAM

ANKIT KOHLI
CHIEF DATA SCIENTIST

Ankit Kohli is Data Science Lead in the space of AI, Machine Learning and Big Data helping organizations across globe in enabling the application of Advanced analytics. With over a decade of his professional experience, he is the lead in data sciences at D Cube Analytics. Prior to this he has worked in data sciences business engagements at Absolutdata, EXL and Cognizant (MarketRX) across industries implementing analytical frameworks to business strategies to augment revenue streams for the businesses.

DHEERAJ KATHURIA
CONSULTANT

Dheeraj is a Consultant at D Cube's India office, has 6.5+ years of Industry experience in Data Analytics. He has analytics and data science experience across various industries like Pharma, Retail, FMCG, Automobile and Digital OTT platforms

OUR AGENDA TODAY

- Click on the corresponding boxes for navigating to their respective sections

dcube™
DATA DRIVEN DECISIONS

Proprietary and Confidential

Data Sources Available

Claims Data

Contains:
- Cost-related information
- Data of insured population
- Diagnosis & treatment info relevant for reimbursement

Sales Data
- Captures physician and account level sales data

Promotion Data
- Promotion carried out and response data

Call Activity Data
- Call Message data

Apps, Social Media
- Captured from live health tracking
- Real-time data of patient health diagnostics
- Captures patient sentiments shared online

Key Attributes Captured

Data Resources

Patient Attributes
- Patient demographics
- Geographic information
- Diagnosis history

Provider Attributes
- Physician specialties
- Provider affiliations
- Provider demographics

MCM & Call Activity
- Digital promotion
- Promotion response
- Influencer data
- Speaker program

Payer Attributes
- Reimbursement and copay
- Payor plan coverage
- Drug accessibility

Sales Data
- Monthly/Weekly prescription level data at prescriber and account level

Innovative Use Cases
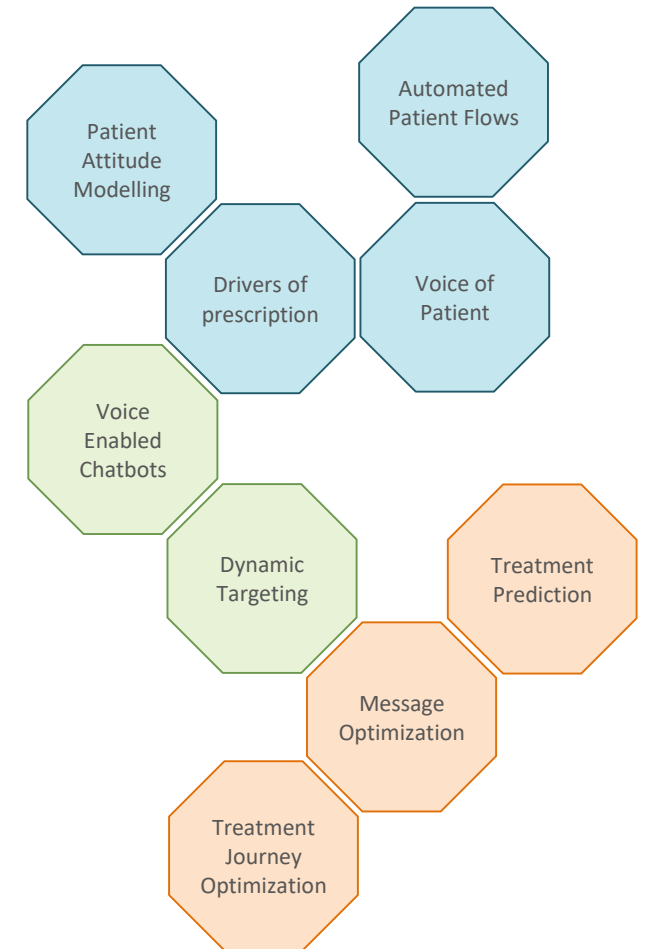
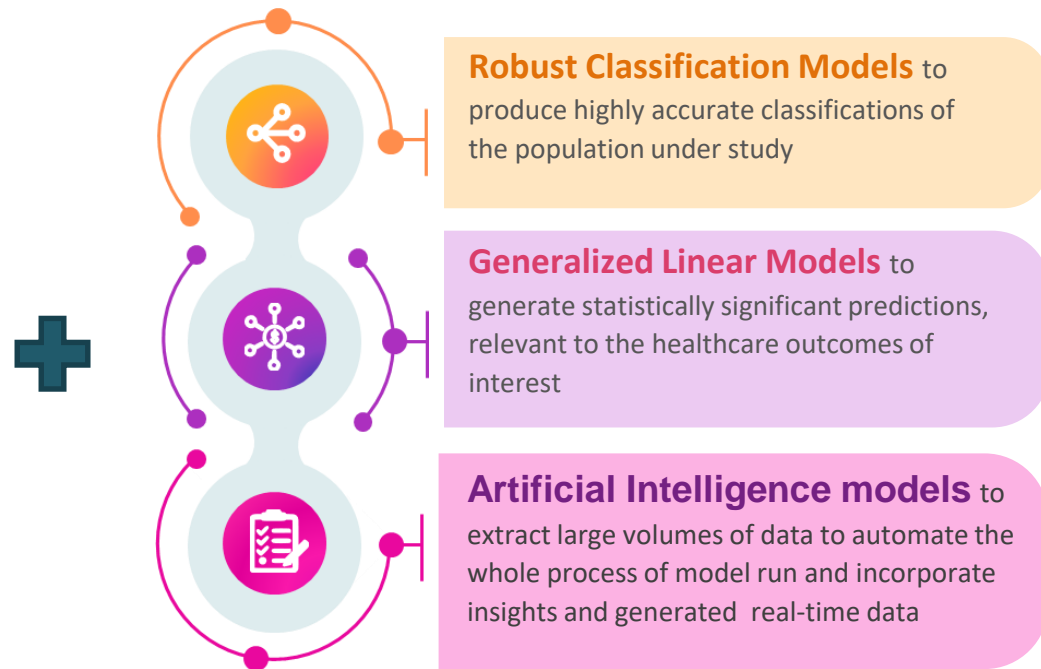- Patient Attitude Modelling
- Automated Patient Flows
- Drivers of prescription
- Voice of Patient
- Voice Enabled Chatbots
- Dynamic Targeting
- Treatment Prediction
- Message Optimization
- Treatment Journey Optimization

Data Sources

Marketing data

Data Sources

Sales, Affiliation, HCP demographics etc

Payer/Claims. APLD, HER/EMR etc

Advanced Analytical Techniques

Robust Classification Models to produce highly accurate classifications of the population under study

Generalized Linear Models to generate statistically significant predictions, relevant to the healthcare outcomes of interest

Artificial Intelligence models to extract large volumes of data to automate the whole process of model run and incorporate insights and generated real-time data

Data Science Use-cases

- Early Adoption Targets Prediction
- MCM Targeting
- Dynamic Targeting
- Segmentation and Profiling
- Loyalty Prediction
- HCP Influencer Mapping
- Channel & Content Optimization
- Payer-Provider Index
- Patient Flow App
- Message Optimization through Rep Notes
- Treatment Prediction
- Voice Enabled Chat Bot

Accelerators

Baseline Business Rules

Ready-to-use standardized business rules and KPI definitions that can be used for the majority of the analysis

Re-usable Code Modules

Reusable R and Python scripts that mirror the baseline business rules and KPI definitions that enable quick-starts on the projects
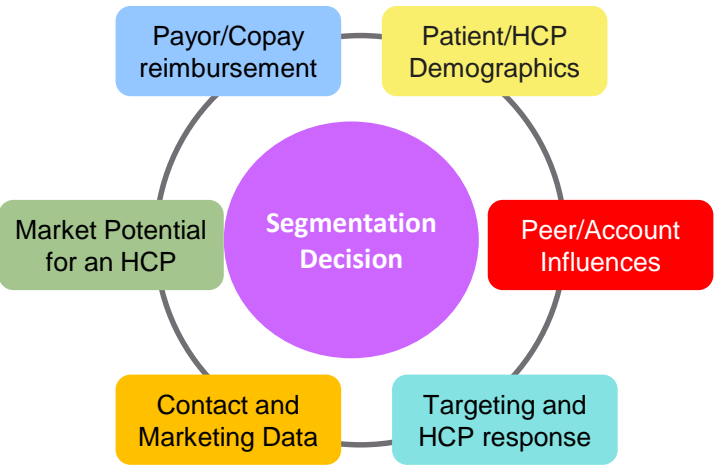
Feature Libraries

Ready to use feature library with relevant features designed over time with numerous engagement delivered across the customers
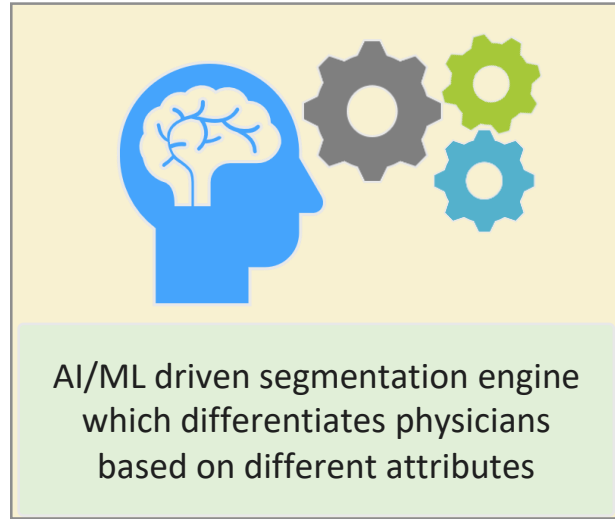
dcube™
DATA DRIVEN DECISIONS

SEGMENTATION ANALYSIS IS ONE SUCH ANALYSIS THAT LEVERAGES AVAILABLE DATA AND ADVANCED CLASSIFICATION TECHNIQUES TO IDENTIFY FUTURE PRESCRIPTION BEHAVIOR
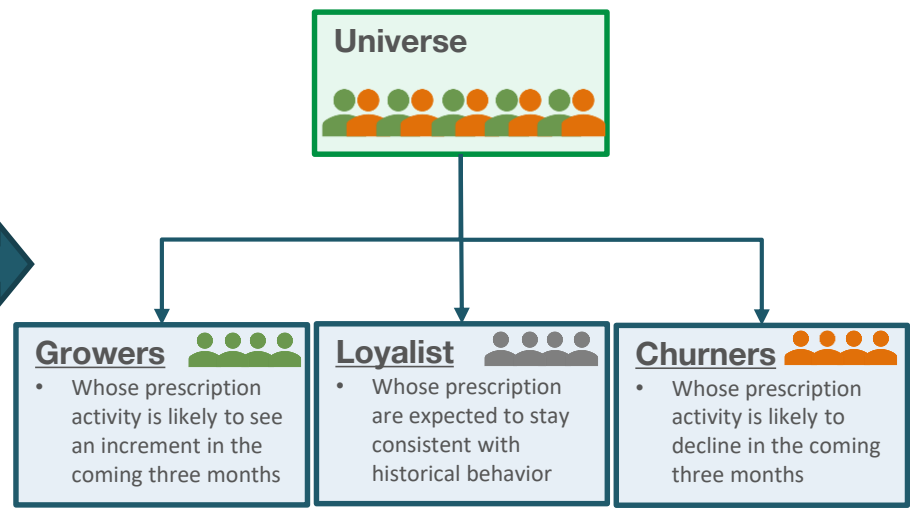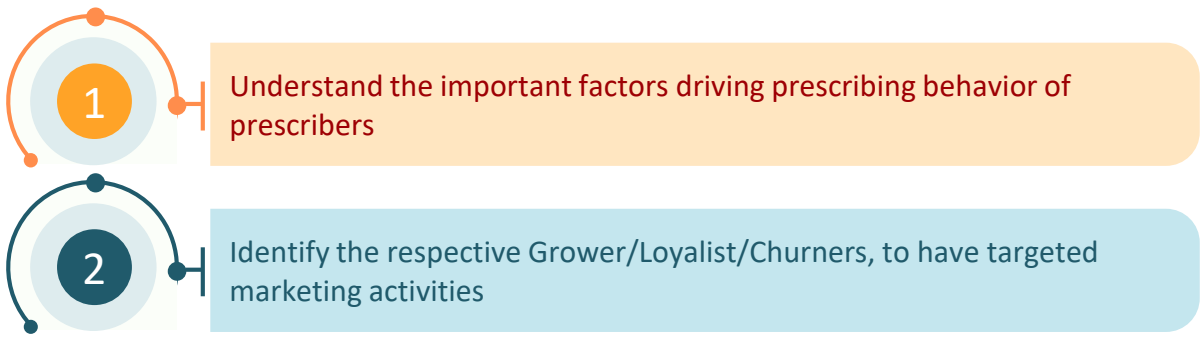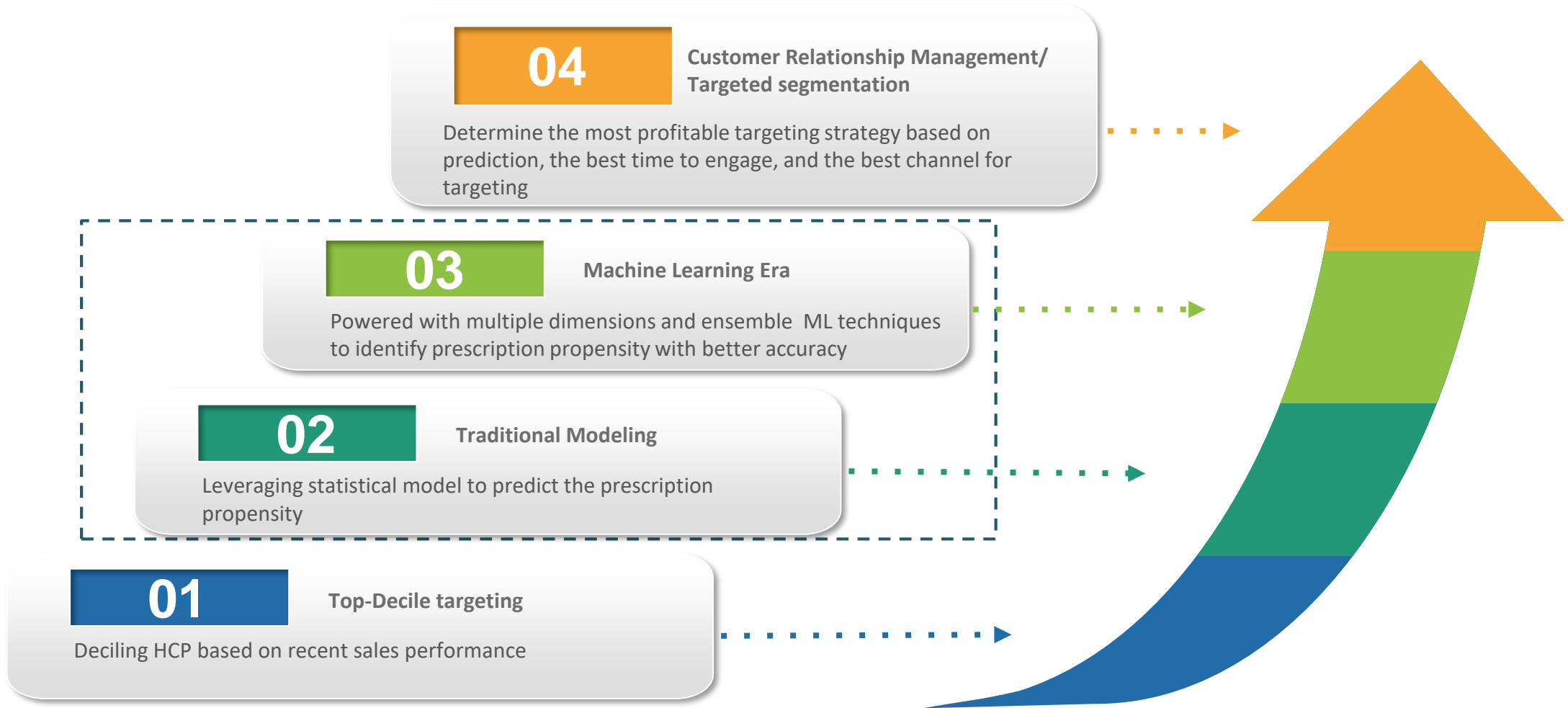
Factors Driving Segmentation

Payor/Copay reimbursement

Patient/HCP Demographics

Market Potential for an HCP

Segmentation Decision

Peer/Account Influences

Contact and Marketing Data

Targeting and HCP response

Prescription Growth Prediction Engine

AI/ML driven segmentation engine which differentiates physicians based on different attributes

Analytical Outputs

Universe

Growers
- Whose prescription activity is likely to see an increment in the coming three months

Loyalist
- Whose prescription are expected to stay consistent with historical behavior

Churners
- Whose prescription activity is likely to decline in the coming three months

Outcomes Targeted

1 Understand the important factors driving prescribing behavior of prescribers

2 Identify the respective Grower/Loyalist/Churners, to have targeted marketing activities

3 Effective use of marketing budgets and increase profitability

4 Basis the identified drivers and the favorable segment profiles, one can design the relevant targeting strategies

dcube™
DATA DRIVEN DECISIONS

04 Customer Relationship Management/ Targeted segmentation

Determine the most profitable targeting strategy based on prediction, the best time to engage, and the best channel for targeting

03 Machine Learning Era

Powered with multiple dimensions and ensemble ML techniques to identify prescription propensity with better accuracy

02 Traditional Modeling

Leveraging statistical model to predict the prescription propensity

01 Top-Decile targeting

Deciling HCP based on recent sales performance

dcube
DATA DRIVEN DECISIONS

THIS SEGMENTATION STUDY ANSWERS THE FOLLOWING KEY QUESTIONS

Key questions

Who? — Who is likely to decline/grow/be loyal to my brand in the next 3 months?

How? — How important is that physician?

What? — What needs to be done to prevent decline or keep growth momentum?
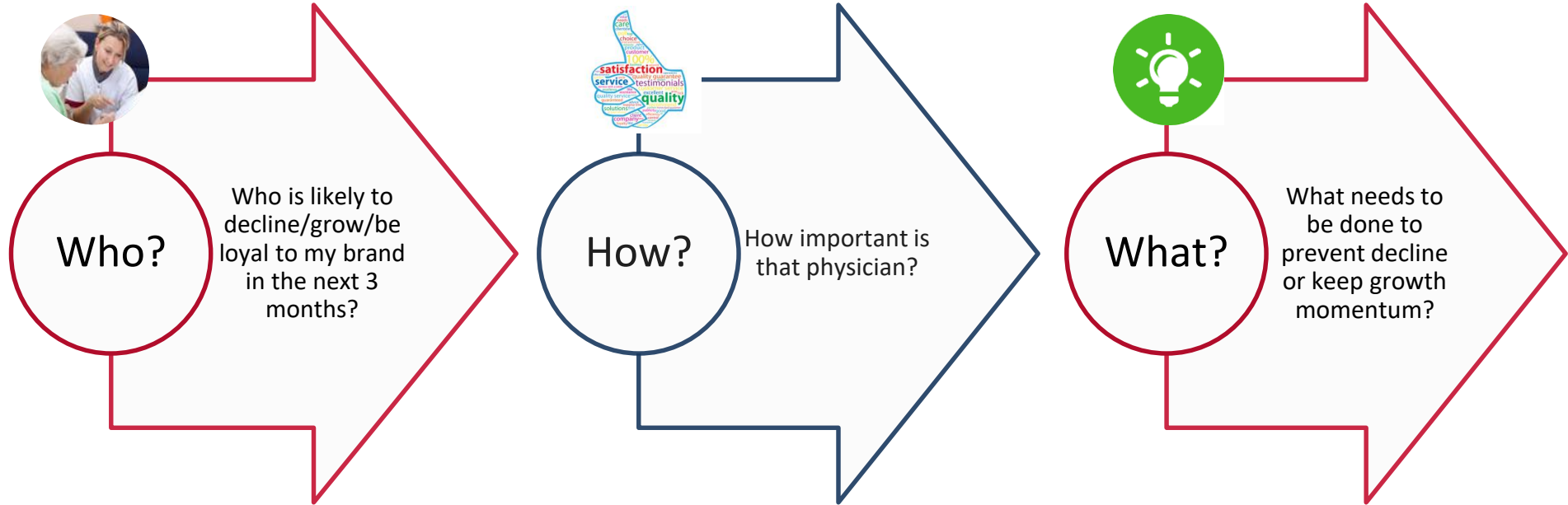
Analytic Solutions

- HCP Rx decline/growth/loyalty prediction model
- Future predicted RX Score and other KPI's like NBRx, claims, #patients, etc.
- Model Explanatory Packages Like LIME & Profiling Segments

Outcome

- Rank ordering of physicians based on predicted Rx growth behavior
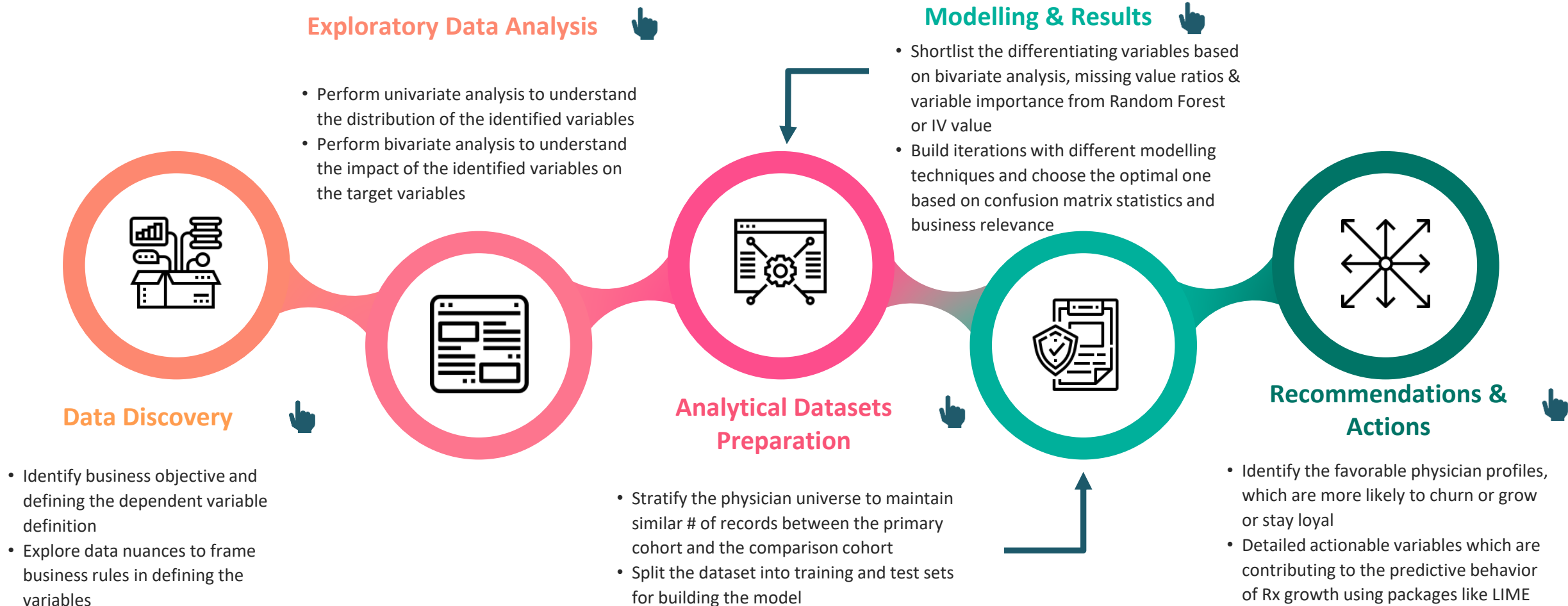- Physician potential (patient volume) measure
- Physician segments to drive actionable decisions

Analysis Objective: Understand the most important factors driving prescription activity of a prescriber

Exploratory Data Analysis

- Perform univariate analysis to understand the distribution of the identified variables
- Perform bivariate analysis to understand the impact of the identified variables on the target variables

Modelling & Results

- Shortlist the differentiating variables based on bivariate analysis, missing value ratios & variable importance from Random Forest or IV value
- Build iterations with different modelling techniques and choose the optimal one based on confusion matrix statistics and business relevance

Data Discovery

- Identify business objective and defining the dependent variable definition
- Explore data nuances to frame business rules in defining the variables

Analytical Datasets Preparation

- Stratify the physician universe to maintain similar # of records between the primary cohort and the comparison cohort
- Split the dataset into training and test sets for building the model

Recommendations & Actions

- Identify the favorable physician profiles, which are more likely to churn or grow or stay loyal
- Detailed actionable variables which are contributing to the predictive behavior of Rx growth using packages like LIME

dcube™
DATA DRIVEN DECISIONS

EVALUATED RELEVANT DATA SOURCES AVAILABLE WITH OUR CUSTOMER FOR MODEL DEVELOPMENT

Physician related

Affiliations

Physician - Professional affiliations, common hospital affiliations

Census / demographic information for physician

Any zip level or physician demographic information, e.g, age, years of experience, Marital status, etc.

Payor related

Health Claims data

Managed market claims data(Rx and Mx)

HCP Co-pay card utilization

Co-pay cards and free samples utilization data

Sales related

Prescriber past sales (Xponent/ Weekly + Specialty data)

Monthly/ Weekly prescription activity at the physician level along with channel

Internal Sales alignment

Semester alignment for past 2 yrs.

Internal Sales - Detailing data

Detailing info including message & physician detailing notes

Call message data

Purpose of the call

Marketing related

HCP Tactics data

Speaker Programs (remote/ in-person) with # of attendees, Journals with # of circulations, HCP Online - Search & Display with # of Clicks, Impressions, Other MCM with # of visits and # of engaged visits, Product Theatres with # of attendees and events

Meeting and events

Physician meetings and events data, KOL influence data

DTC marketing data

Data containing DTC GRP and marketing spend related information

Seepage Ratio: A measure of change in prescription activity that we considered for event rate calculation

Latency Period

Past Avg 6 month(Observation Window)

Future Avg 3 months(Prediction Window)

Latency Period: Two months lag was introduced so that the marketing team can devise targeting strategy for individual segment

Explored different combination of P3-F3, P4-F4, P4-F2, P6-F3 etc and finally arrived at P6-F3 combination

Prescriber Universe

Step 1: **Based on metrics involving past prescription activity for an HCP**

For whole Prescriber Universe, set of filters used were
- Market Share of Drug A NRx
- Market Share of Drug A TRx
- Volume Change in Nrx
- Volume Change in Trx

It was an OR condition.

Step 2: Defining Growers/Loyalist/Churners

After selecting the metric for defining the segments, we select respective cutoff for individual segment to arrive at our final dependent variable for modelling purpose.
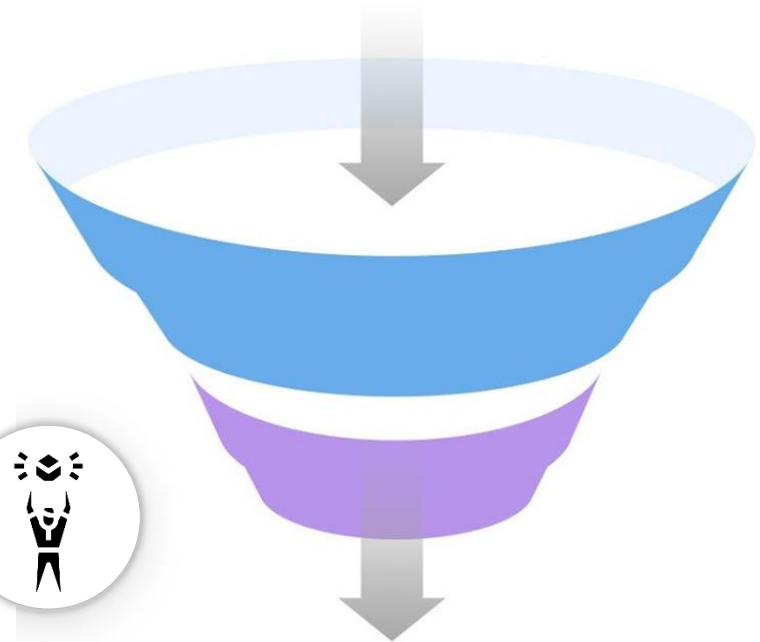
ILLUSTRATIVE

1. Data Source

Data Sources
- Captures point of care health information
- Claims related information: payor type, insurance type etc
- Drug and procedural information
- Physician demographic
- Copay and free sample data
- Provider specialty
- Speaker program data
- Marketing level data

2. Univariate Analysis

Estimated the richness of data source for all the variables and fill rate for all variables and check for transformation if necessary
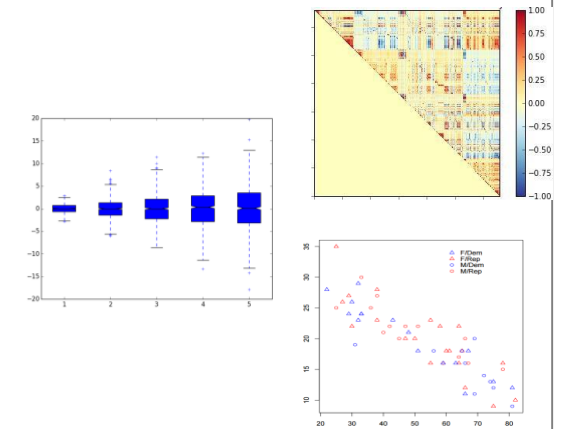
- **Central Tendency: Mean, Median, Mode , Fill rate etc**

- **Standard Deviation, IQR etc**

- **Histograms, Box Plots to check the normality of the variables**
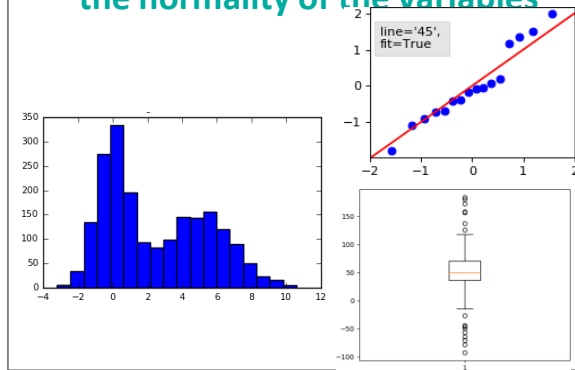
3. Bivariate analysis

Performed Bivariate analysis to analyse the effect of variables on the cohorts

- **Cross tabulation plots**
- **Correlation matrices plots**
- **Scatter plots**

1 Master Data Creation

Extract and integrate all the custom variables into one table with the outcome of the event

Robust Tracking:

- Ensure continuous coverage of physician-level data
- Coverage across Pharmacy and Medical Claims both
- Ensure unique value for each variable
- Coverage of physician-level data across various data sources

Integrating the Datasets:

- Combine attributes from multiple tables
- Apply deduplication rules to create final dataset

2 Feature Engineering

Convert the data variables into modelling input features by applying required transformations

One-Hot Encoding:

- Convert categorical variables into numerical variables
- Explore various ratio, % changes , average and several count variables for activity, and marketing response variables

Custom Variables Creation:

- Create custom attributes from the raw attributes, more relevant for the analysis
- Final data to be at each outcome level

3 Final Analytical Dataset

Perform required treatments to generate the final dataset to run our models on

Pre-processing:

- Using a final list of features, stratify the model to give equal weightage to target cohort and comparison cohort
- Based on stratification, split the dataset into training and test dataset for trying different ML algorithms

Variable Importance:

- Run the model on the training dataset and validate on the test dataset
- Obtain variable importance and accuracy metrics. Use performance tuning to arrive at the best fit model

d cube
DATA DRIVEN DECISIONS

VARIABLES SELECTION

Derived variable creation

Variable selection for profiling

From ~300 basic variables, ~1800 derived variables were created to measure the historical occurrence of activity

To capture the magnitude of the activity

Count of activity (e.g. calls, claims, etc.) in the past 1 month

Monthly average of activity in the past 3 months

Monthly average of activity in the past 6 months

To capture the trend of the activity in recent months

Ratio of activity in the past 3 months v/s past 6 months

Ratio of activity in the past 3 months v/s. past previous 3 months

The difference of activity in the past 3 months v/s past previous 3 months

Segregated all the variables into 5 categories – Sales, Marketing, Calls, Attributes & Payor

Calculated Information Value (IV) of all the variables

Selected the variables with IV of greater than equal to 0.1 across the categories

Identified ~70 variables from the 5 categories which are representative of the category based on insights

METHODOLOGY FOLLOWED

1. Modeling dataset creation

- Created ~1800 variables based on a hypothesis
- Split the dataset into development (80%) and validation (20%) with non-overlapping physicians in each data set
- Stratify the samples appropriately

2. Variable selection

- Correlation checks: Out of the multiple variations for a root variable like past 2 months average, past 6 month average, etc., P3 and P3byP6* had the highest correlation with the dependent variable
- Multicollinearity: Remove multicollinearity (based on VIF)

3. Model development

Traditional method

Logistic Regression

- Ran multiple iterations of logistic regression to identify variables based on their significance and impact

GLMM

- Classified the variables as potential fixed & random effects and fitted the model to obtain drivers

✔ Rheum target list suppressed the importance of other variables in logistic regression
✔ Therefore, GLMM was chosen as it captures the impact of variables appropriately across different groups

Machine Learning

Random Forest

- Filtered the variables basis their importance and impact on the target variable
- Tuned the parameters to achieve the best performing model

XGBoost

- Shortlisted variables based on variable importance and partial dependency plots
- Turned the hyper-parameters

✔ XGBoost model gave a better performance than Random Forest
✔ XGBoost model showed significantly improved performance than GLMM

Ensemble

GLMM + XGBoost

- Took weighted average of the predictions from each modelling technique
- Built a predictive model with prediction of each model used as predictors

✔ Ensemble showed only *marginal improvement* in performance compared to XGBoost
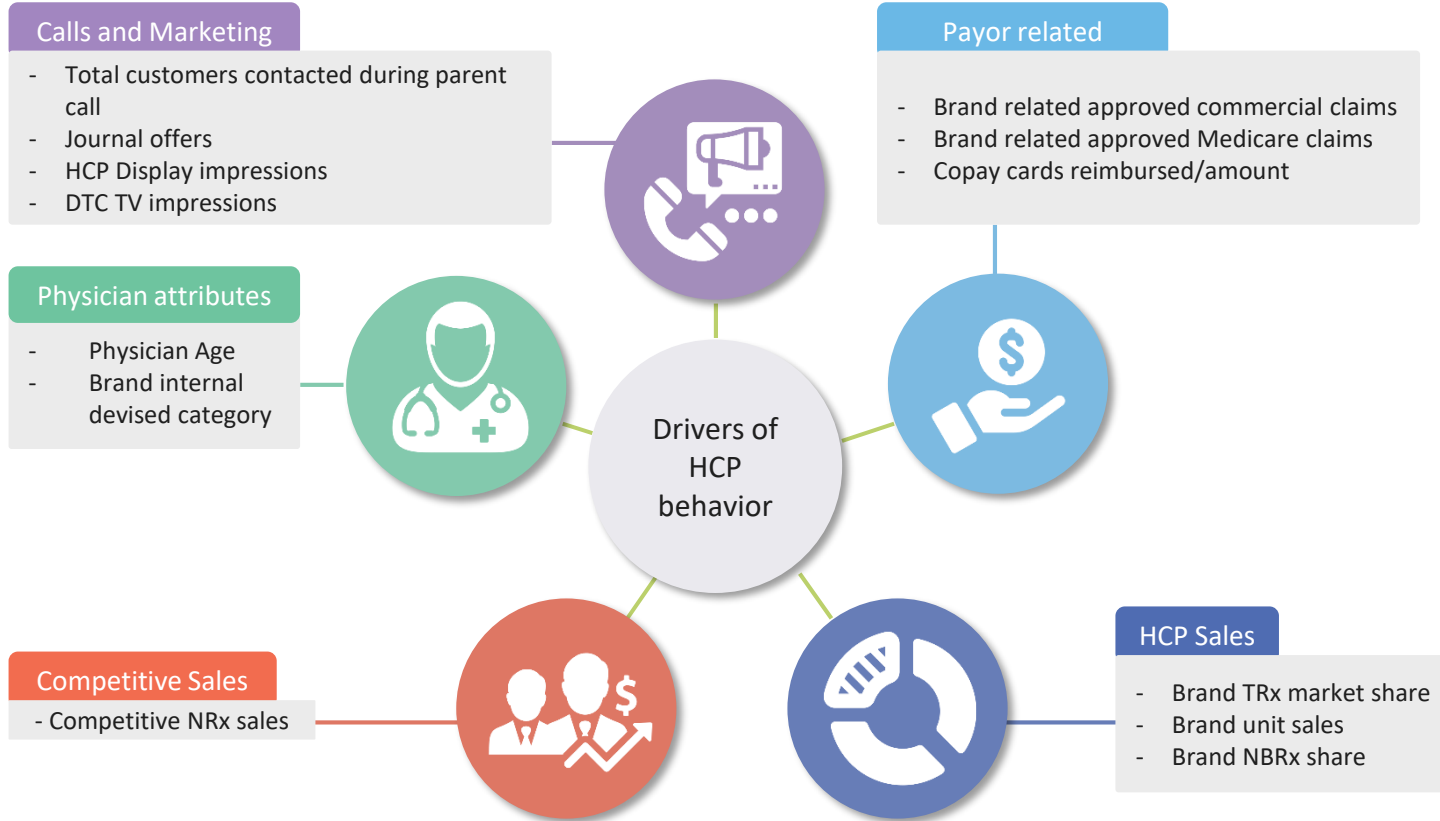
4. Model validation

- Hold-out sample validation: Assessed performance of the model i.e. AUC (Area under the curve), KS statistics & capture rate on validation data
- Cross-fold validation: Assessed performance of the model using a 5-fold cross-validation method for robustness
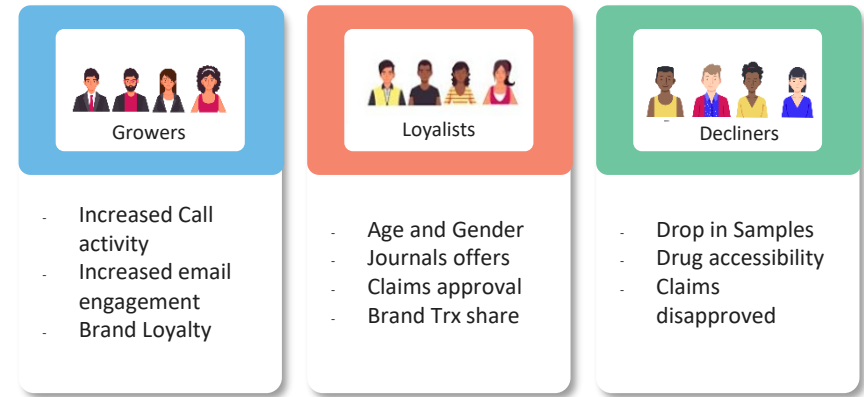
*Ratio of past 3 month average to past 6 month average

Drivers of Engagement

Segmented Personas

Calls and Marketing
- Total customers contacted during parent call
- Journal offers
- HCP Display impressions
- DTC TV impressions

Physician attributes
- Physician Age
- Brand internal devised category

Competitive Sales
- Competitive NRx sales

Payor related
- Brand related approved commercial claims
- Brand related approved Medicare claims
- Copay cards reimbursed/amount

Drivers of HCP behavior

HCP Sales
- Brand TRx market share
- Brand unit sales
- Brand NBRx share

Growers
- Increased Call activity
- Increased email engagement
- Brand Loyalty

Loyalists
- Age and Gender
- Journals offers
- Claims approval
- Brand Trx share

Decliners
- Drop in Samples
- Drug accessibility
- Claims disapproved

Personalized drivers of prescription

Predicted Behavior | **Drivers of prescription**

Grower → Increased Call Activity | Attended Speaker Event | Influencer Network

Decliner → No Recent Connect | Drop in Samples | Reduced Patients

Loyalist → Stable Call Activity | Influencer Network | Claim Approval

Q&A
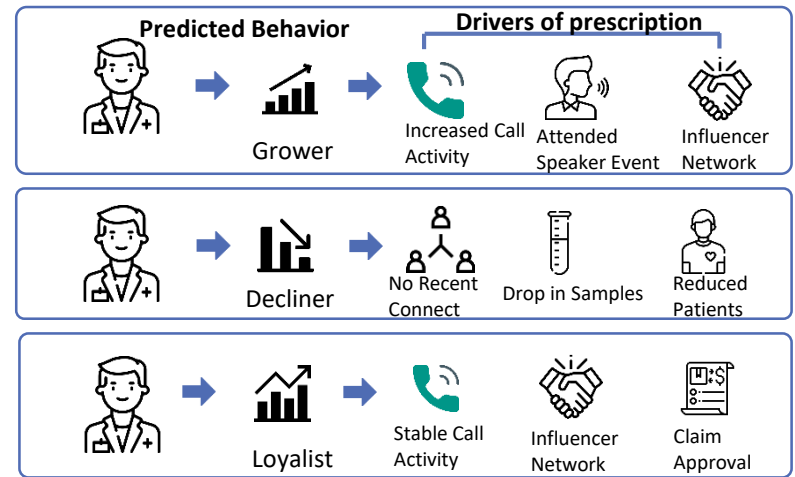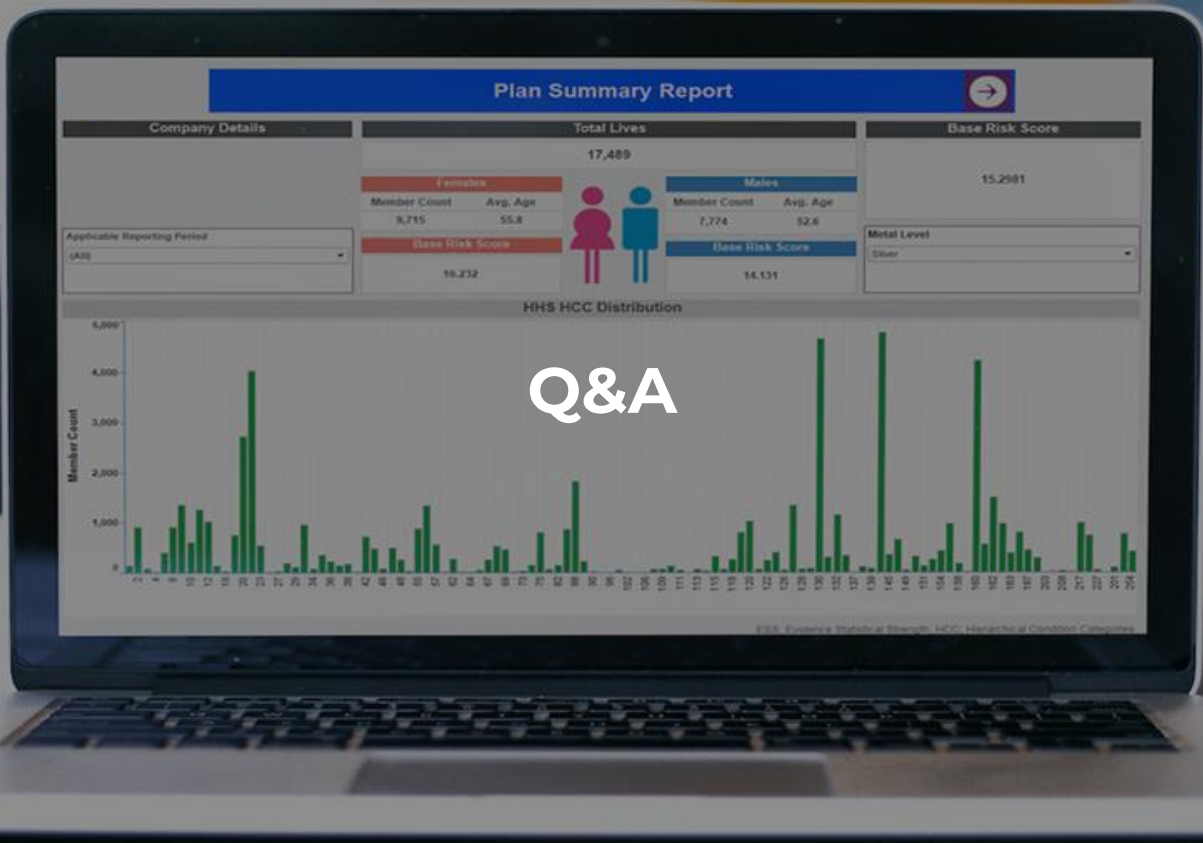
READY TO TEST DRIVE
THE NEW PARADIGM?

Contact

Email
info@dcubeanalytics.com

Contact

Phone
US : +1 847.807.4996

US
Office
D Cube Analytics Inc.1320 Tower
Road, Schaumburg, Illinois 60173,
USA