Data Governance and Analytics for Healthcare

Anji Technologies

Forecasting Demand & Optimizing B2B Transactions

<u>Business Problem</u>

Large variations in patient inflow resulted in under-utilization of resources impacting revenue. Untimely transfer of patients to a specialized care resulted in a lower level of customer satisfaction while not able to admit new patients. Large revenue cycles especially on pharmacy claims in certain specialized areas led to an increase in unearned revenues

<u>IT Solution provided</u>

❑ Provided a data governance solution to standardize master data across various departments

❑ Provided visualizations and adhoc analysis capability to visualize and identify problem areas proactively. Identifying problem areas will help stream line B2B transactions

❑ Provided Machine learning models to predict length of stay and outcomes to effectively manage demand and supply

<u>Results achieved</u>

❑ Reduced variation in patient inflow resulted in optimization of resource utilization and increased revenue by 3% - 5% *year over year

❑ Transfer of patients to appropriate health care facility in timely manner resulted in improvement in customer satisfaction by 2%* year over year

❑ Streamlined the reimbursement process for pharmacy related claims improving the rate of reimbursement by 4-6% *in specialties of Circulatory diseases and neoplasms

*varies by healthcare provider

Business Intelligence Architecture
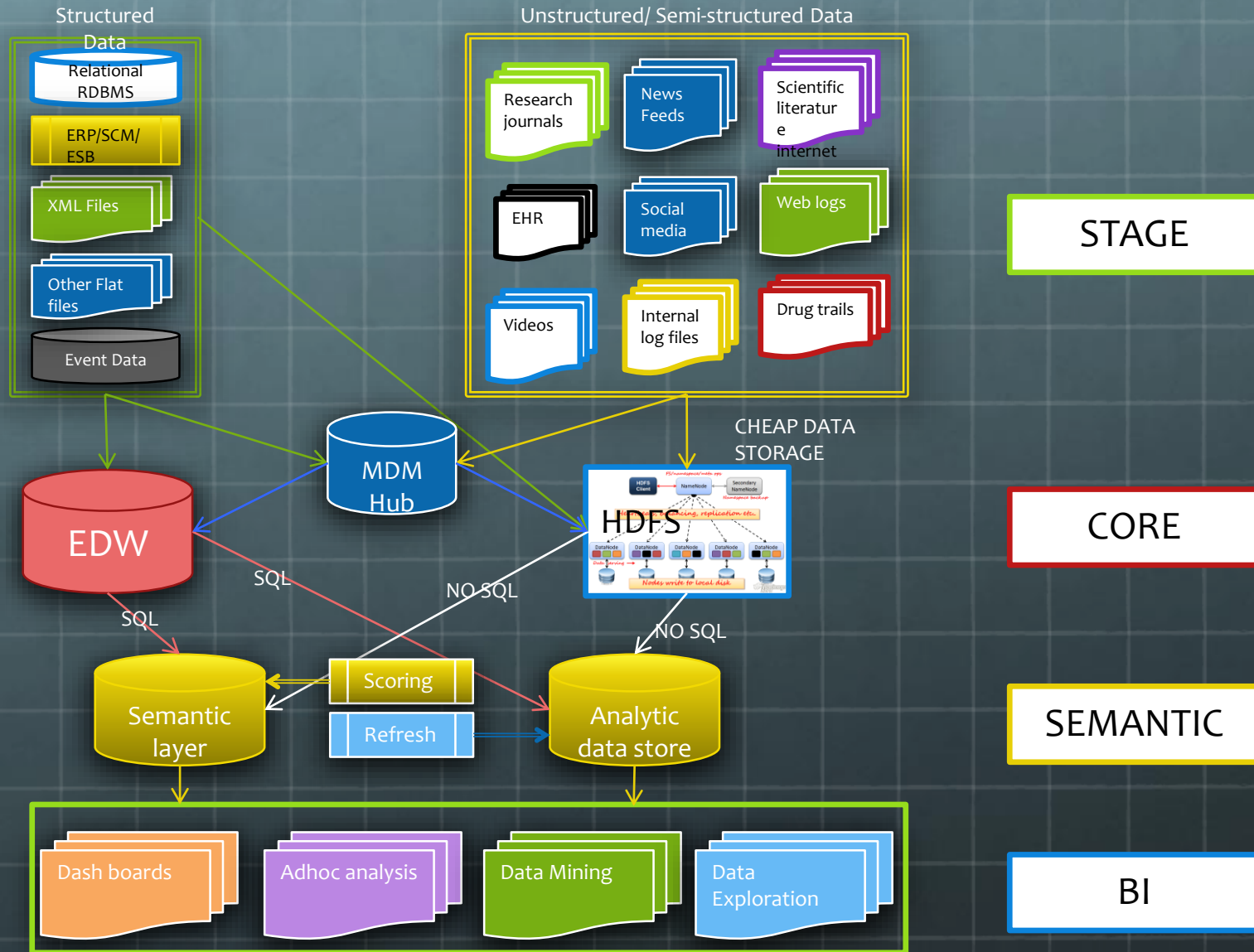
BI Strategy for Data Analytics, Integrated Insights and Advanced Visualization

Recommendations for creating best in class BI solution

- Create appropriate hierarchies for performing drill downs

- Conform the master data across all departments utilizing an MDM solution

- Set up a performant semantic layer from the core layer using techniques such as summarization at database level or utilizing an OLAP strategy

- Choose a traditional reporting tool for dashboards, an analytical tool for data exploration and insights and a data mining tool for building models and scoring the input data

MDM Architecture
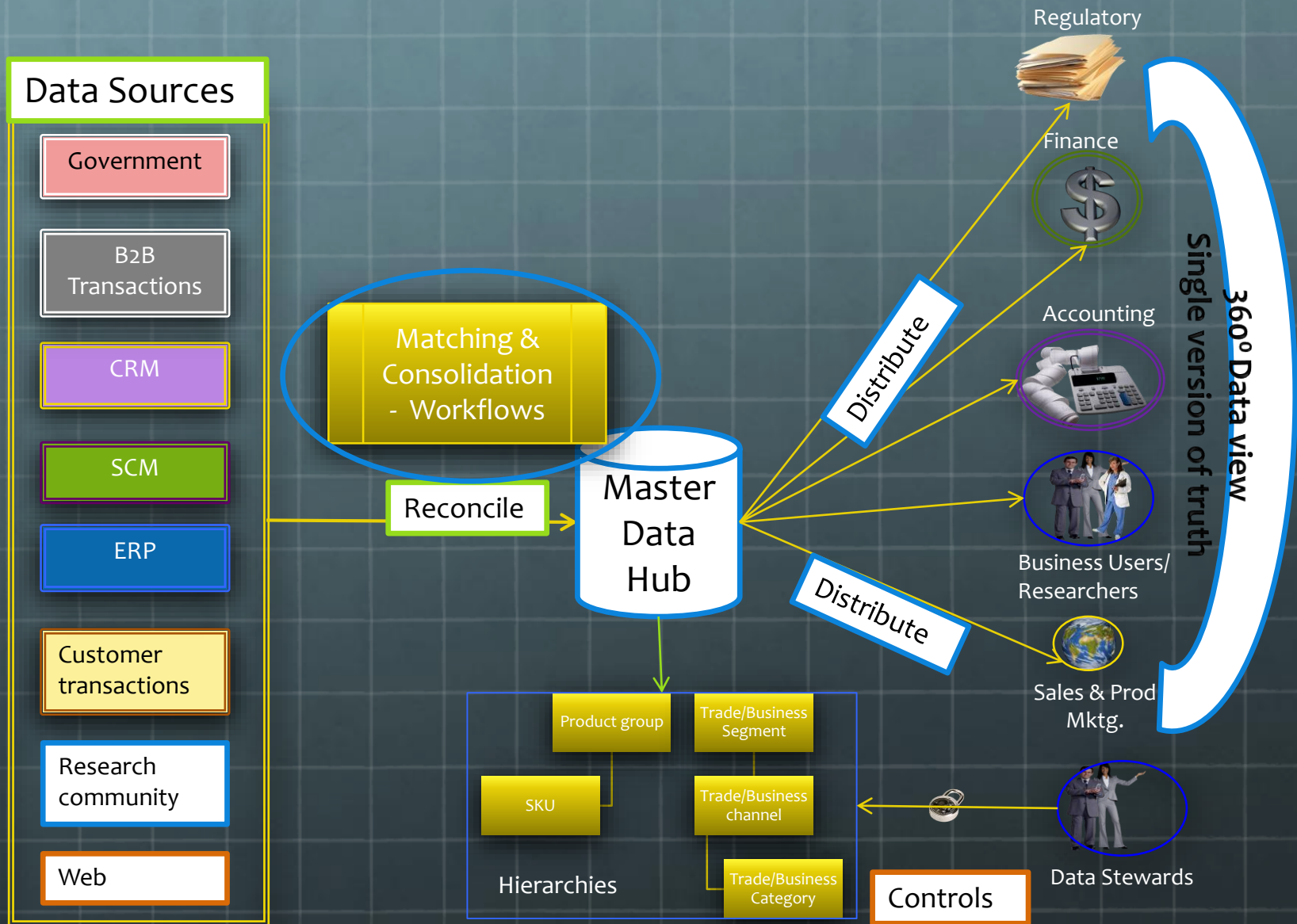
Master Data Management - Implementation

o **Achieved a 360° view of the customer**

o **Created a single version of a business entity and attributes**

o **Provided centralized workflows that track flow of master data from source to semantic layer**

o **Provided a process to create hierarchies and associate appropriate base elements**

o **Created a centralized rules repository for association of various master data elements between subject areas and transactions**

o **Provided appropriate mechanism for approvals to change business rules or associations between various master data elements**

Business Intelligence and Analytics

Data is loaded into an enterprise data warehouse and Hadoop. Conformed data into semantic layer utilizing SQL and No-SQL to create the following BI components:

1. Dashboards

2. Advanced visualizations

3. Adhoc analysis to see detailed data

Capabilities Provided

1. Ability to perform text filtering

2. Ability to associate data elements and deduce relationships on the fly

3. Ability to traverse hierarchy within visualization

4. Ability to filter the data within the report page

5. Ability to see the details of the data on demand with a click of a button

6. Ability to tie visualizations together

7. Integrated BI capabilities with HDFS for advanced data mining and exploration

Basic Dashboards

Basic dashboards usually are distributed to decision makers across business domains to gauge pre-determined outcomes. Ex: In this case visualize pharmacy dollars by source of admission for a given time period usually given as a filter

A Check box filter embedded on the dashboard

Dashboards take basic forms such as a tabular or a cross table form. However many reporting tools offer following visualizations put together in a dashboard

➢ Bar chart
➢ Line chart
➢ Combination chart
➢ Pie chart
➢ Scatterplot
➢ Boxplot
➢ Map plot

Advanced Visualizations – Tree Map with Hierarchies

Average length of stay and average pharmacy dollars as % total dollars plotted by quarter for various disease groups and sub groups

Controls bottom panel
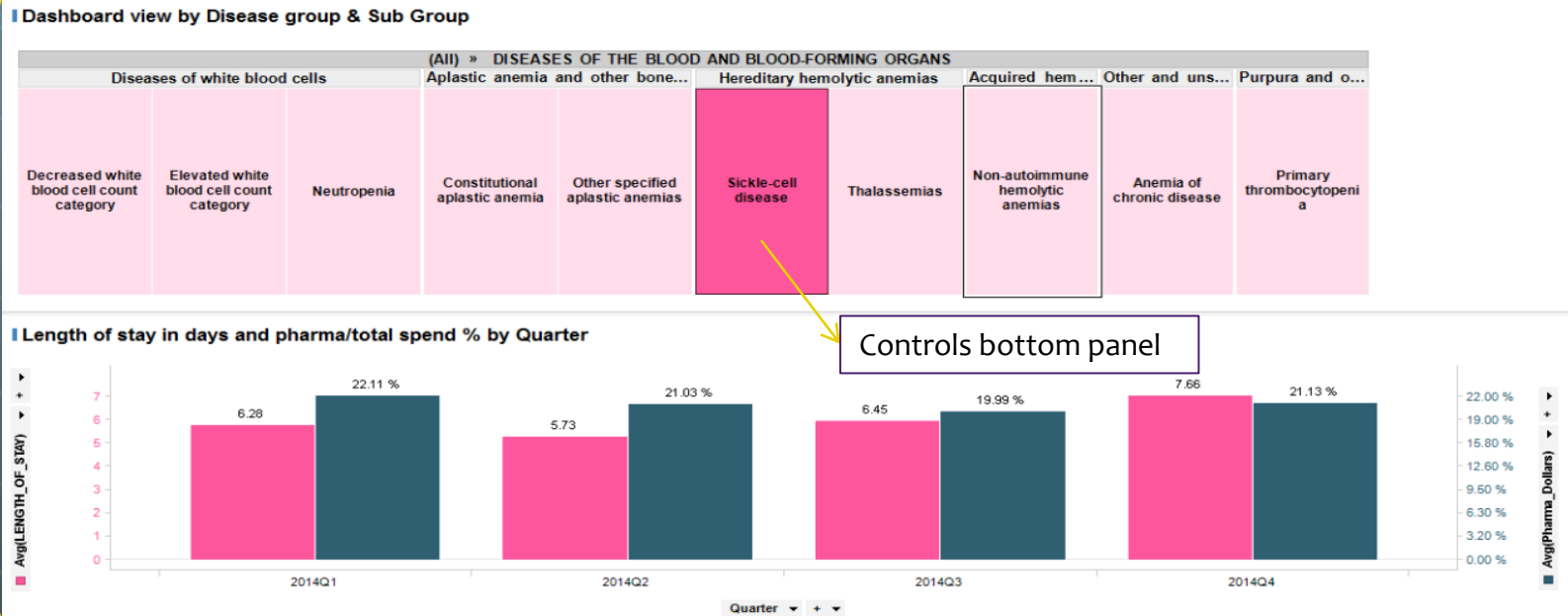
The above chart provides insights into various disease groups and subgroups for which the length of the hospital stay can be assessed real time with a click of a button on a tree map. Similar visualizations such as heat maps provide real time display of activity based on a master data element. Above chart also shows hierarchies defined and visualized such as Diseases of blood to Hereditary hemolytic anemias to sickle-cell disease for which metrics are shown in the bottom chart by quarter.

Adhoc Data Analysis

In order to understand the underlying data more adhoc analysis need to be performed. To narrow down the data appropriate selections & filtering should be performed. Selections, searching text values and filters can be done graphically.
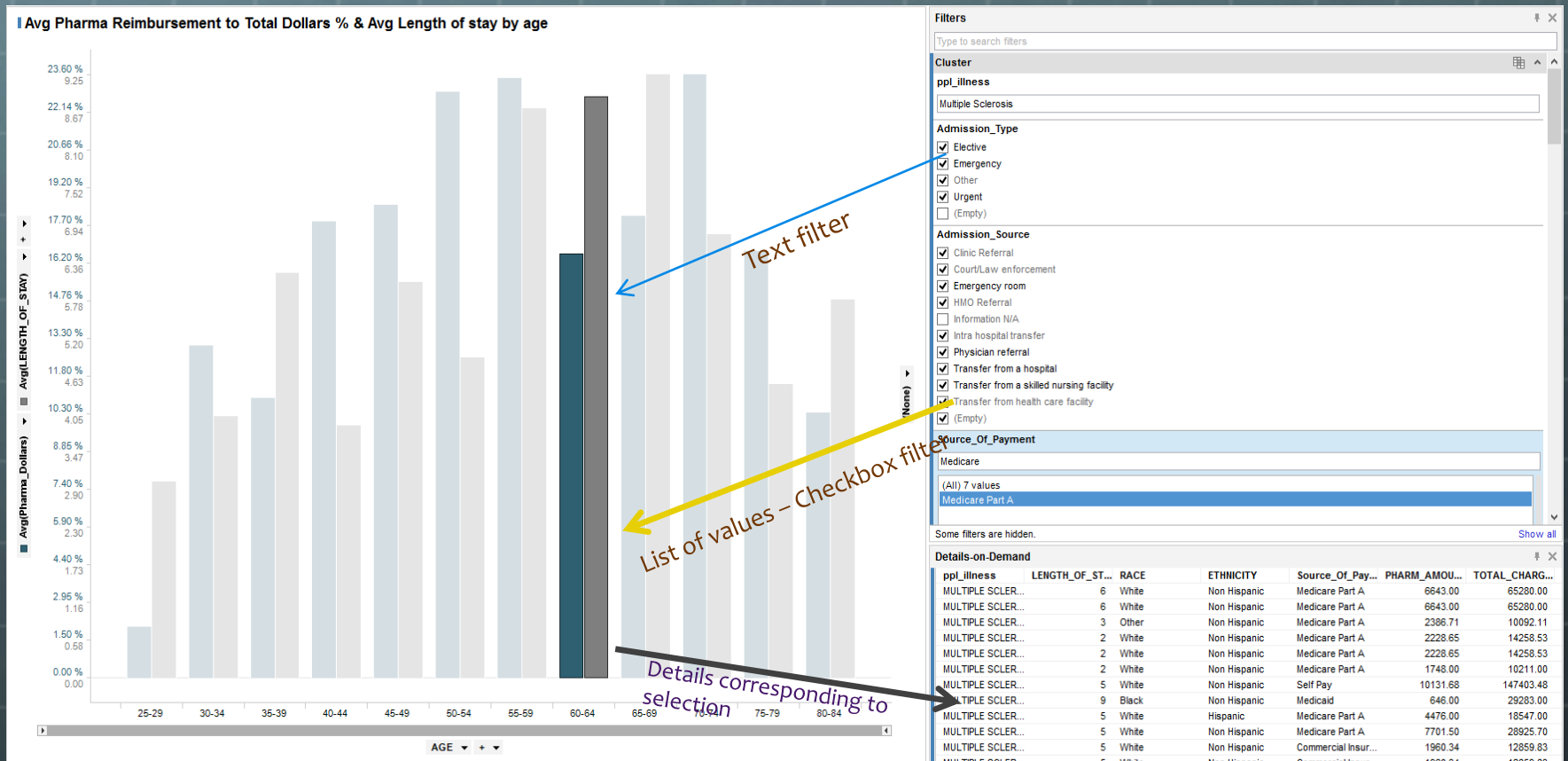
Integrated Insights

- **Provided integrated insights on the patients**

- **Provided ability to drill down on disease groups**

- **Provided the ability to slice and dice**

- **Provided the ability to perform adhoc reporting**

- **Created cutting edge visualizations using Tree maps and heat maps**

- **Provided capability to perform data exploration using associations**

Data Mining

o Petabytes of patient data such as demographics, vital signs, diagnosis & treatment performed, medications used and the notes and voice recording data from the health care professionals is stored in a Hadoop cluster and queried using Pig & Hive SQL

o Challenge is preparing the data for data mining as most of the mining algorithms require normalized data and a representative sample

o Training data and test data should be maintained separately. Also models should be setup with a learning rate to make them adjust for the new data patterns.

o Moreover joining the transactional data with conformed dimensions is a key to prepare categorical variables acceptable for all the departments.
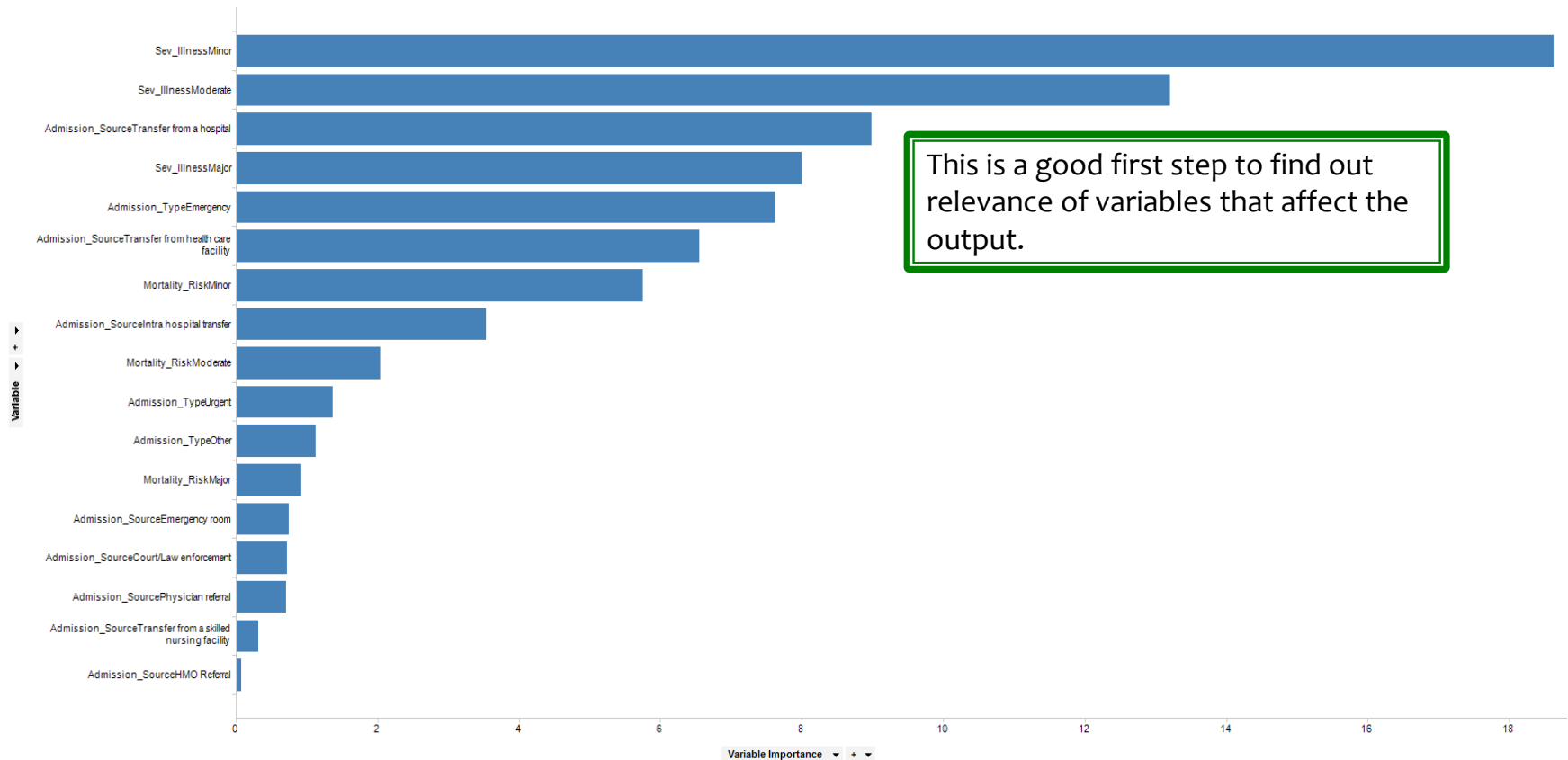
Classification Modeling – Logistic Regression

Model Name: MyModel
Model Type: Logistic Regression
Model: LENGTH_OF_STAY == "2" ~ Admission_Type + Admission_Source + Sev_Illness + Mortality_Risk

Variable Importance

This is a good first step to find out relevance of variables that affect the output.

K-means Clustering

Clear well defined clusters represent a classification of particular interest when analyzed with other data elements they provide opportunities/ meaningful insights. In this case it turned out to be reimbursement practices by insurance companies in certain class of drugs used for specific illnesses such as neoplasms & diseases of circulatory system

Cluster 1 generates maximum Pharmacy dollars/Total dollars and Cluster 6 generates minimum pharmacy dollars/ Total spend. Study practices causing the anomaly.
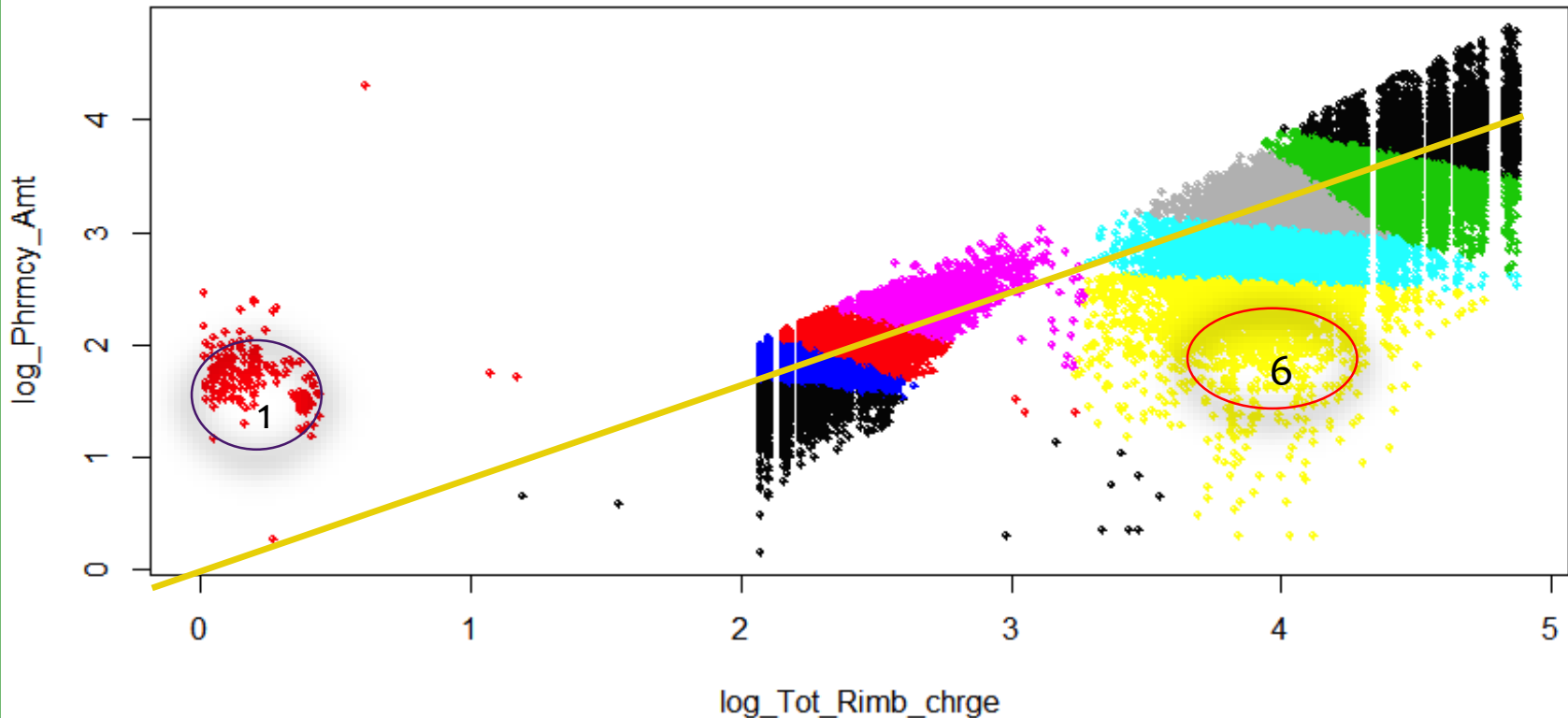
Artificial Neural Network - MLP

Machine learning algorithms such as neural networks can be trained to provide an output based on some input variables. MLP models can be provided with a learning rate to adjust ongoing new input conditions and adjust outputs accordingly. The trained model should be thoroughly tested with test data while reducing over fit.
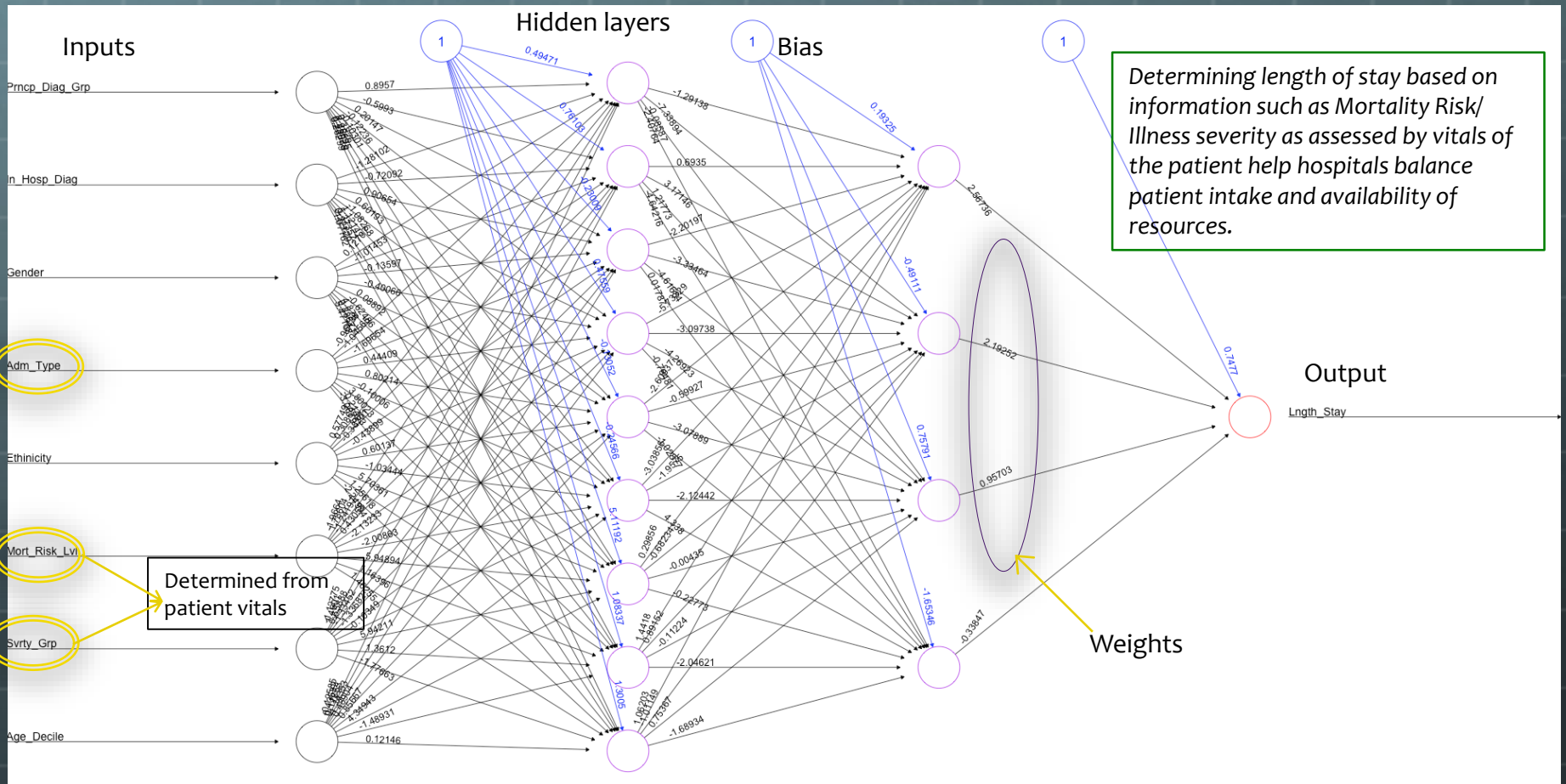
Neural Networks contd..

Assessing the outcomes such as discharging to a specialized facility after a time period is really helpful to be proactive transferring such patients to a specialized care. Typical examples of this case is transfer to a psychiatric care facility at the right time or transferring to a specialized unit to better the outcome.
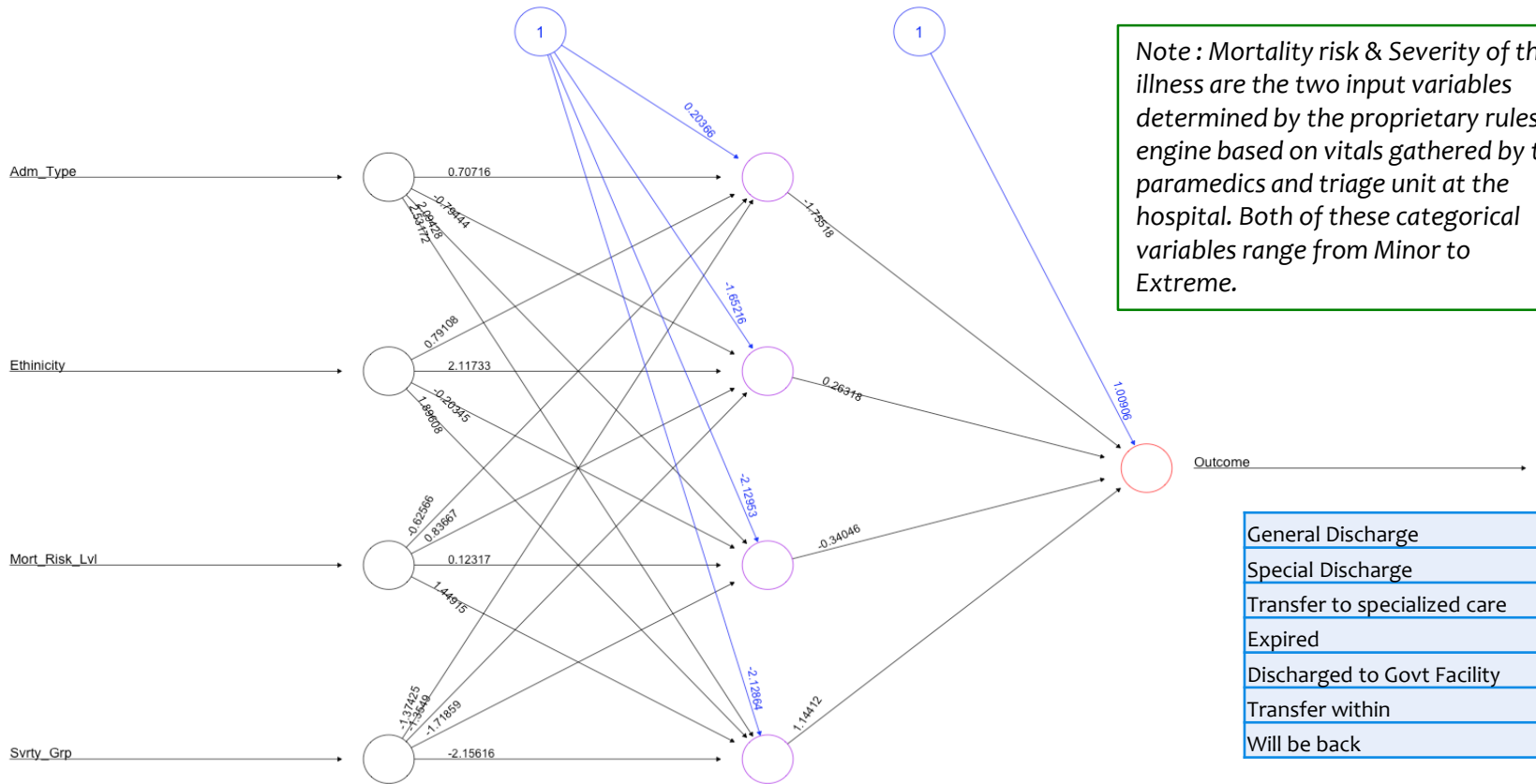
Note : Mortality risk & Severity of the illness are the two input variables determined by the proprietary rules engine based on vitals gathered by the paramedics and triage unit at the hospital. Both of these categorical variables range from Minor to Extreme.

| General Discharge |
| Special Discharge |
| Transfer to specialized care |
| Expired |
| Discharged to Govt Facility |
| Transfer within |
| Will be back |

Predictive Models and Scoring

- Created analytic data store to store information generated from the models

- Provided the capability to perform scoring of the models

- Tied the scoring information back with the semantic layer

- Provided ability for the models to interface with real time data and provide the outcomes on the real time basis

Capabilities - Tools and Technologies Used

Hadoop implementations - utilizing MAPR, Cloudera, ASTER Data & Horton HDP

MAPREDUCE – Java, Python, ASTER SQL MAP REDUCE

EDW – Oracle RAC, Teradata,DB2

Data Mining – SAS, R, SPSS, MATLAB, MAHOUT

RDBMS – SQLServer, DB2, Teradata, Oracle

ETL/ELT – Informatica, Abinitio, Datastage, SQOOP, FLUME

MDM – Informatica MDM, Abinitio, IBM Initiate, Oracle Hyperion EPM

BI – Tableau, Microstrategy, Cognos BI, Business Objects, TIBCO Spotfire & QLIKVIEW

NOSQL/ NOSQL databases – HIVE, PIG, VERTICA, HBASE, MONGODB, Cassandra

Automation & Monitoring – PERL, Autosys, MAESTRO, CONTROLM, AMBARI, OOZIE