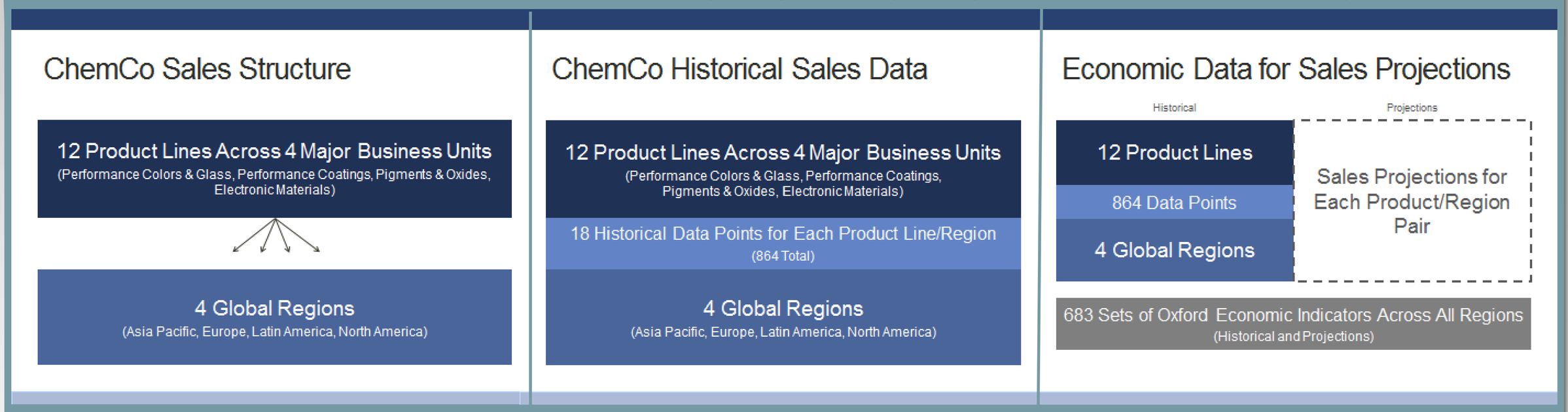
Capgemini Challenge: Beat Our Projections

Machine Learning Prevails over Regression

Background:

- Capgemini's client, ChemCo, sells chemical coatings to a variety of industrial clients
- Common protocol for sales projections is to ask the sales force to estimate but sales managers are rewarded for beating estimates. This led to sandbagging of estimates.
- ChemCo needs a model that is derived from external data in order to project sales through 2018.
- If sales are correlated with economics, a predictive model leveraging macroeconomic data may be possible.
- Capgemini established a linear regression model for ChemCo.

Task: Predict future sales for ChemCo by improving on previous models with given data.



Methodology:

1. Try to duplicate Capgemini regression models for understanding
2. Utilize machine learning random forest models for improvement

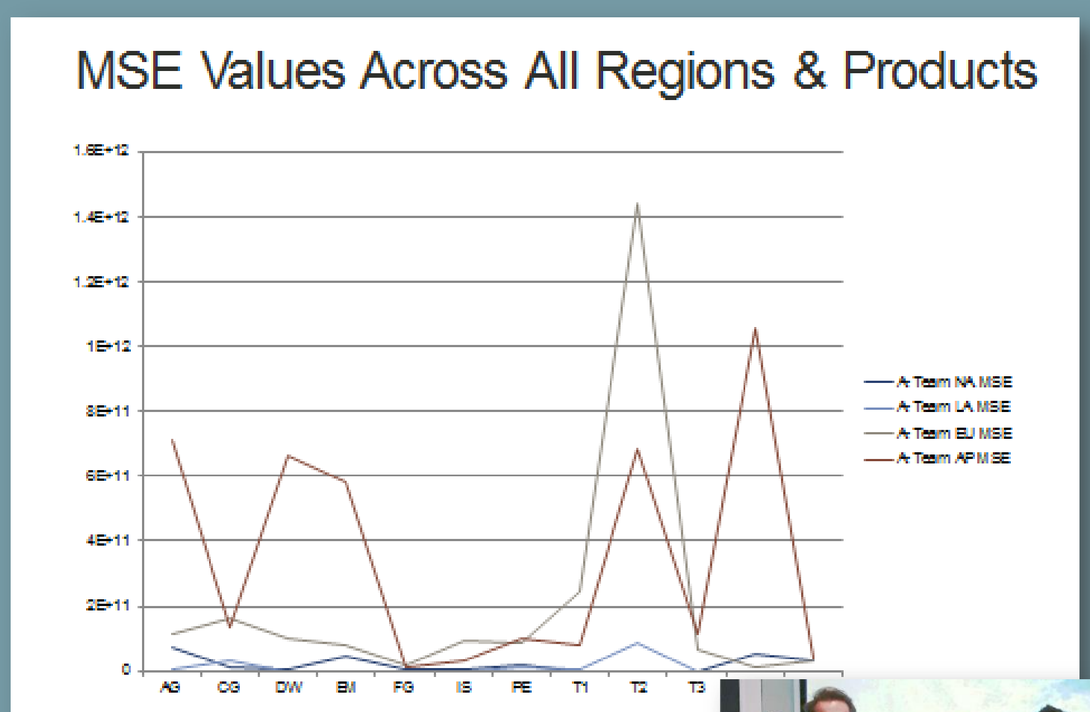
Why Random Forest?

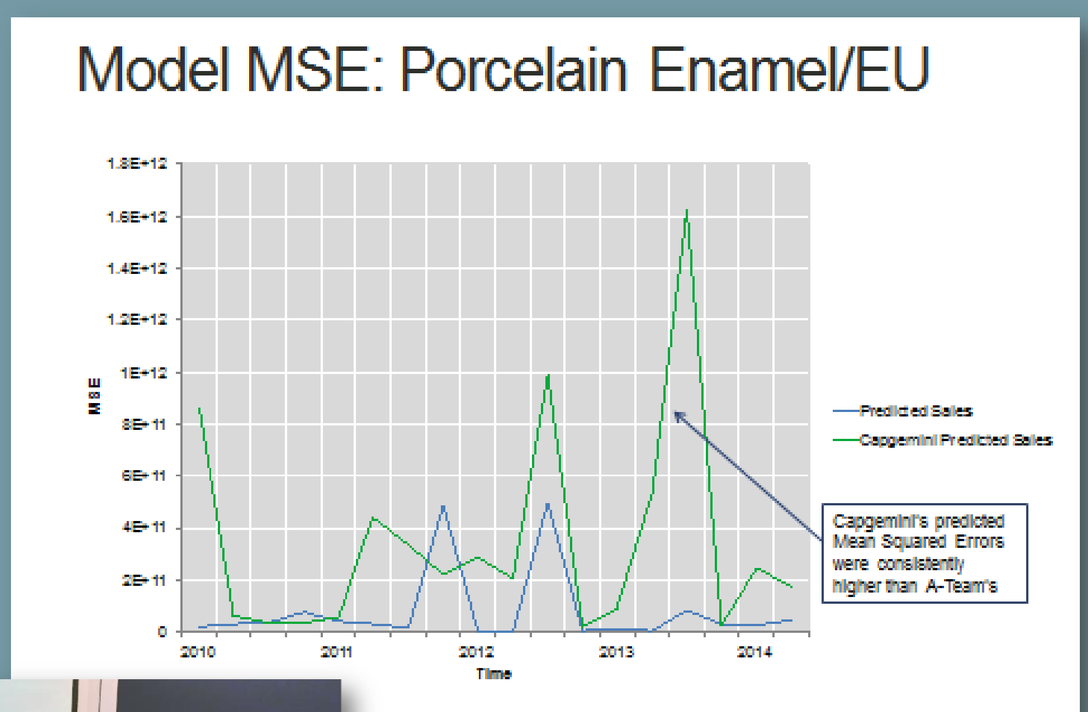
- **"Small n, large p" problems,** i.e. data sets with a small number of observations but large number of features
- **Correlated predictor variables**, as was the case with the multiple economic indicator variables
- Focus not only on prediction, but also **on assessing variable importance**

Challenges We Faced:

- Only 18 quarters of historical data per product/region made it hard to train and test
- Possible serial correlation

Results:





Analytics Lab 2014 | Capgemini Sponsored | Mentor: Frank MacCrory
Team: Phatty Arbuckle | Amitava Bhattacharya | Andrew Flamang | Akshay Singh