



PEARSON BUSINESS ANALYTICS SERIES

Product Analytics

Applied Data Science Techniques for
Actionable Consumer Insights

Joanne Rodrigues



Product Analytics

Product Analytics: Applied Data Science Techniques for Actionable Consumer Insights

Table of Contents

Cover

Half Title

Title Page

Copyright Page

Contents

Preface

Acknowledgments

About the Author

Part I: Qualitative Methodology

1 Data in Action: A Model of a Dinner Party

1.1 The User Data Disruption

1.1.1 Dont Leave the Users out of the Model

1.1.2 The Junior Analyst

1.1.3 The Opposite of the Misguided Analyst: The Data Guru

1.2 A Model of a Dinner Party

1.2.1 Why Are Social Processes Difficult to Analyze?

1.2.2 A Party Is a Process

1.2.3 A Party Is an Open System

1.2.4 A Great Party Is Hard to Define

1.2.5 Party Guests Motives and Opinions Are Often Unknown

1.2.6 A Party Presents a Variable Search Problem

1.2.7 The Real Secret to a Great Party Is Elusive

Table of Contents

1.3 Whats Unique about User Data?

- 1.3.1 Human Behavior Is a Process, Not a Problem
- 1.3.2 No Clear and Defined Outcomes
- 1.3.3 Social Systems Have Rampant Problems of Incomplete Information
- 1.3.4 Social Systems Consist of Millions of Potential Behaviors
- 1.3.5 Social Systems Are Often Open Systems
- 1.3.6 Inferring Causation Is Almost Impossible

1.4 Why Does Causation Matter?

1.5 Actionable Insights

2 Building a Theory of the Social Universe

2.1 Building a Theory

- 2.1.1 Wont Fancy Algorithms Solve All Our Problems?
- 2.1.2 The Pervasive (and Generally Useless) One-Off Fact
- 2.1.3 The Art of the Typology
- 2.1.4 The Project Design Process: Theory Building
- 2.1.5 Steps to a Good Theory
- 2.1.6 Description: Questions and Goals
- 2.1.7 Analytical: Theory and Concepts
- 2.1.8 Qualities of a Good Theory

2.2 Conceptualization and Measurement

- 2.2.1 Conceptualization
- 2.2.2 Measurement
- 2.2.3 Hypothesis Generation

2.3 Theories from a Web Product

- 2.3.1 User Type Purchasing Model
- 2.3.2 Feed Algorithm Model
- 2.3.3 Middle School Dance Model

2.4 Actionable Insights

3 The Coveted Goalpost: How to Change Human Behavior

3.1 Understanding Actionable Insight

3.2 Its All about Changing Your Behavior

- 3.2.1 Is It True Behavior Change?

Table of Contents

3.2.2 Quitting Smoking: The Herculean Task of Behavioral Change

3.2.3 Measuring Behavior Change

3.3 A Theory about Human Behavioral Change

3.3.1 Learning

3.3.2 Cognition

3.3.3 Randomized Variable Investment Schedule

3.3.4 Outsized Positive Rewards and Mitigated Losses

3.3.5 Fogg Model of Change

3.3.6 ABA Model of Change

3.4 Change in a Web Product

3.5 What Are Realistic Expectations for Behavioral Change?

3.5.1 What Percentage of Users Will See a Real Change in Our Product?

3.5.2 Are Certain Behaviors Easier to Change?

3.5.3 Behavioral Change Worksheet

3.6 Actionable Insights

Pat II: Basic Statistical Methods

4 Distributions in User Analytics

4.1 Why Are Metrics Important?

4.1.1 Statistical Tools for Metric Development

4.1.2 Distributions

4.1.3 Exploring a Distribution

4.1.4 Mean, Median, and Mode

4.1.5 Variance

4.1.6 Sampling

4.1.7 Other Measures

4.1.8 The Exponential Distribution

4.1.9 Bivariate Distribution

4.2 Actionable Insights

5 Retained? Metric Creation and Interpretation

5.1 Period, Age, and Cohort

5.1.1 Period

5.1.2 Age

Table of Contents

- 5.1.3 Cohort
- 5.1.4 Lexis Diagram
- 5.1.5 Period versus Cohort?
- 5.1.6 Cohort Retention
- 5.1.7 Period Retention
- 5.1.8 Standardization

5.2 Metric Development

- 5.2.1 Conceptualization
- 5.2.2 Acquisition
- 5.2.3 Ratio-Based Metrics
- 5.2.4 Retention
- 5.2.5 Engagement
- 5.2.6 Revenue
- 5.2.7 Progression Ratios

5.3 Actionable Insights

6 Why Are My Users Leaving? The Ins and Outs of A/B Testing

- 6.1 An A/B Test
- 6.2 The Curious Case of Free Weekly Events
 - 6.2.1 Spurious Correlation
 - 6.2.2 Selection Bias
- 6.3 But Its Correlated
 - 6.3.1 Proportional Comparisons
 - 6.3.2 Linear Correlations
 - 6.3.3 Nonlinear Relationships
- 6.4 Why Randomness?
- 6.5 The Nuts and Bolts of an A/B Test
 - 6.5.1 A/B Test Process
 - 6.5.2 The Setup
 - 6.5.3 Randomization
 - 6.5.4 Hypothesis Testing
 - 6.5.5 Power Analysis
- 6.6 Pitfalls in A/B testing

Table of Contents

- 6.6.1 General Pitfalls
- 6.6.2 No Randomness
- 6.6.3 Differing Patterns Between Groups
- 6.6.4 Differing Patterns in Long- and Short-Run Effects

6.7 Actionable Insights

Part III: Predictive Methods

7 Modeling the User Space: k-Means and PCA

- 7.1 What Is a Model?
- 7.2 Clustering Techniques
 - 7.2.1 Segmenting Users, Novice Users, and Unsupervised Learning
- 7.3 Actionable Insights

8 Predicting User Behavior: Regression, Decision Trees, and Support Vector Machines

- 8.1 Predictive Inference
- 8.2 Much Ado about Prediction?
 - 8.2.1 Applications of Predictive Algorithms
 - 8.2.2 Prediction in Behavioral Contexts Is Rarely Just Prediction
- 8.3 Predictive Modeling
 - 8.3.1 Simple Explanation: Methods
- 8.4 Validation of Supervised Learning Models
 - 8.4.1 k-Fold Cross-Validation
 - 8.4.2 Leave-One-Out Cross-Validation
 - 8.4.3 Precision, Recall, and the F1-Score
- 8.5 Actionable Insights

Appendix

9 Forecasting Population Changes in Product: Demographic Projections

- 9.1 Why Should We Spend Time on the Product Life Cycle?
- 9.2 Birth, Death, and the Full Life Cycle
- 9.3 Different Models of Retention

Table of Contents

9.3.1 The Transition Matrix

9.3.2 Snowmobile Transition Example

9.4 The Art of Population Prediction

9.4.1 Population Projection Example

9.4.2 User Death by a Thousand Cuts

9.4.3 Exponential Growth Example

9.5 Actionable Insights

Part IV: Causal Inference Methods

10 In Pursuit of the Experiment: Natural Experiments and Difference-in-Difference Modeling

10.1 Why Causal Inference?

10.2 Causal Inference versus Prediction

10.3 When A/B Testing Doesn't Work

10.3.1 Broader Social Phenomena

10.3.2 Historical Action

10.3.3 No Infrastructure

10.4 Nuts and Bolts of Causal Inference from Real-World Data

10.4.1 Causal Inference Terminology

10.4.2 Natural Experiments

10.4.3 Operationalizing Geographic Space: Difference-in-Difference Modeling

10.5 Actionable Insights

11 In Pursuit of the Experiment, Continued

11.1 Regression Discontinuity

11.1.1 Nuts and Bolts of RD

11.1.2 Potential RD Designs

11.1.3 The Enemy of the Good: Nonrandom Selection at the Cut Point

11.1.4 RD Complexities

11.1.5 Graphing the Data

11.2 Estimating the Causal Effect of Gaining a Badge

11.2.1 Comparing Models

11.2.2 Checking for Selection in Confounding Variables

Table of Contents

11.3 Interrupted Time Series

11.3.1 Simple Regression Analysis

11.3.2 Time-Series Modeling

11.4 Seasonality Decomposition

11.5 Actionable Insights

12 Developing Heuristics in Practice

12.1 Determining Causation from Real-World Data

12.2 Statistical Matching

12.2.1 Basics of Matching

12.2.2 What Features Cause a User to Buy?

12.2.3 Matching Theory

12.3 Problems with Propensity Score Matching

12.3.1 Omitted Variable Bias and Better Matching Methods

12.3.2 No Coverage

12.4 Matching as a Heuristic

12.5 The Best Guess

12.6 Final Thoughts

12.7 Actionable Insights

13 Uplift Modeling

13.1 What Is Uplift?

13.2 Why Uplift?

13.3 Understanding Uplift

13.4 Prediction and Uplift

13.5 Difficulties with Uplift

13.5.1 Lalonde Data Set

13.5.2 Uplift Modeling

13.5.3 Two-Model Approach

13.5.4 Interaction Model

13.5.5 Tree-Based Methods

13.5.6 Random Forests

13.6 Actionable Insights

Table of Contents

Part V: Basic, Predictive, and Causal Inference Methods in R

14 Metrics in R

14.1 Why R?

14.2 R Fundamentals: A Very Basic Introduction to R and Its Setup

14.2.1 The R Language

14.2.2 RStudio

14.2.3 Installing Packages

14.2.4 RMarkdown

14.2.5 Reading Data into R

14.2.6 NAs in R

14.3 Sampling from Distributions in R

14.4 Summary Statistics

14.5 Q-Q Plot

14.6 Calculating Variance and Higher Moments

14.7 Histograms and Binning

14.8 Bivariate Distribution and Correlation

14.8.1 Calculating Metrics

14.9 Parity Progression Ratios

14.10 Summary

15 A/B Testing, Predictive Modeling, and Population Projection in R

15.1 A/B Testing in R

15.1.1 Statistical Testing

15.1.2 Power Analysis

15.2 Clustering

15.2.1 k-Means

15.2.2 Principal Components Analysis

15.3 Predictive Modeling

15.3.1 Linear Regression

15.3.2 Logistic Regression

15.3.3 Decision Trees

15.3.4 Support Vector Machines

Table of Contents

15.3.5 Cross-Validation

15.4 Population Projection

15.4.1 Example 1: User Death by a Thousand Cuts

15.4.2 Example 2: The Exponential Growth Example

15.5 Actionable Insights

16 Regression Discontinuity, Matching, and Uplift in R

16.1 Difference-in-Difference Modeling

16.2 Regression Discontinuity and Time-Series Modeling

16.2.1 Regression Discontinuity

16.2.2 Interrupted Time Series

16.2.3 Seasonality Decomposition

16.3 Statistical Matching

16.3.1 Plotting Balance

16.3.2 Caliper Matching

16.3.3 GenMatch()

16.4 Uplift Modeling

16.4.1 Two-Model Solution

16.4.2 Interaction Model

16.4.3 Causal Conditional Inference Forest and Uplift Forest Models

16.5 Actionable Insights

Appendix

Conclusion

Bibliography

Index