Adding Optimization to Your Analytics Toolbox

Dr. Gwyneth Butera and Dr. Russell Halper

GUROBI OPTIMIZATION

The World's Fastest Solver

end-to-end ANALYTICS

Agenda

- Basics of Optimization and MIP (Mixed Integer Programming)
- Identifying Optimization Problems Within Your Organization
- How MIP Complements Machine Learning

Basics of Optimization and MIP (Mixed Integer Programming)

Dr. Gwyneth Butera

Mathematical Optimization Primer

- What do we mean by mathematical optimization?
- What does "optimal" mean?
- Basic form of a mathematical programming problem (MP)

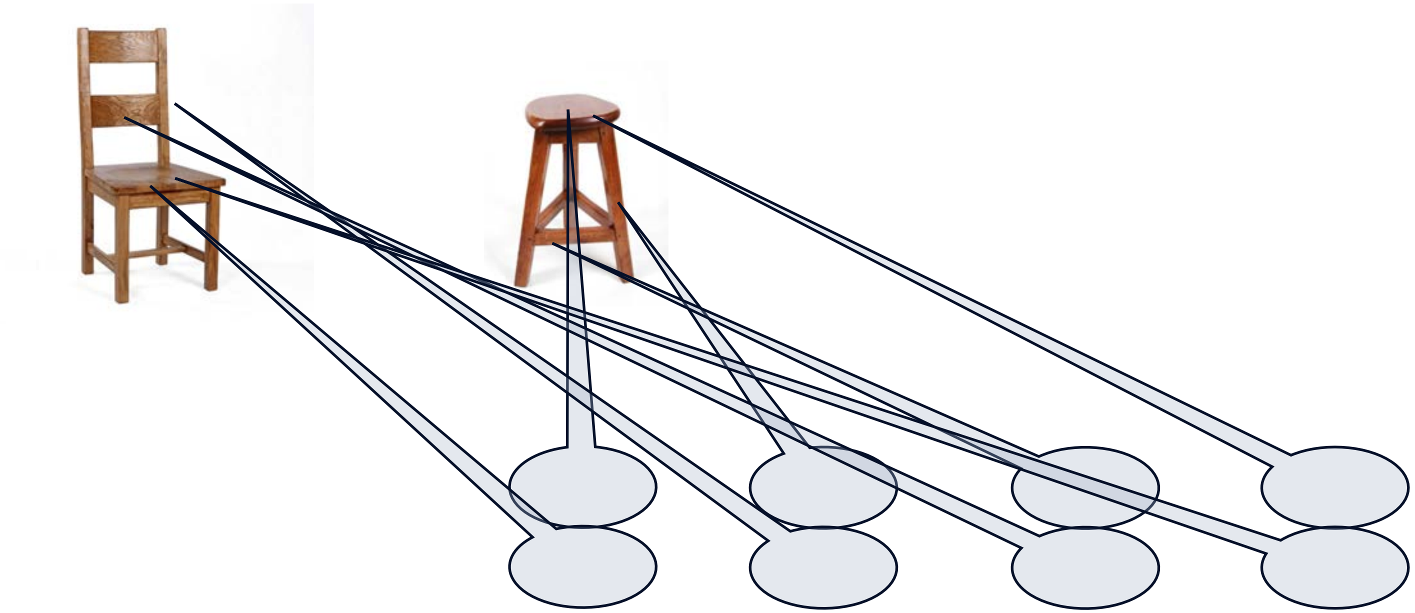
Objective:	minimize $c^T x$
Constraints:	$A x = b$ (linear constraints)
	$l \leq x \leq u$ (bound constraints)
	some or all x_j must take integer values (integrality constraints)

- Many types of MP problems including LP, MIP, QP
- Typical modeling applications

Basic Mixed-Integer Programming Example

"Modular" furniture manufacturing

- b – number of bar stools,
- c – number of chairs.



$$\max_{b,c} \; (10\,b + 11\,c)$$

subject to

$$b + c \leq 10,$$
$$3b + 4c \leq 36,$$
$$3b + 5c \leq 41,$$
$$4b + 2c \leq 35,$$

integer $b, c \geq 0$

Sample Formulation: Sports Scheduling Modeling

- Given a set of teams and time slots
- Produce a schedule for the matches between teams
- Goals
  - Fair and exciting schedule
- Constraints
  - Must obey league rules (*e.g.* number of home/away games)
  - No schedule conflicts (*e.g.* venue restrictions)
  - Potentially lots of others

Sports Scheduling: Basic Model

- Define a set of binary variables X
  - X(i, j, k) = 1 if team i plays team j in week k
  - X(i, j, k) = 0 otherwise

- Translate scheduling rules into linear constraints
  - Each team plays a game each week
    - X(i, j1, 1) + X(i, j2, 1) + X(i, j3, 1) + … = 1
  - Each team must eventually play every other team (in league, in division, …)
    - X(i, j, 1) + X(i, j, 2) + X(i, j, 3) + … = 1

Sports Scheduling: Basic Model

Prohibited games (teams 1 and 4 can't play in week 3)

- X(1,4,3) = 0

Required games (teams 3 and 5 must play in week 7)

- X(3,5,7) = 1

No back-to-back division series (1 and 2 can't play week 3 and week 4)

- X(1, 2, 3) + X(1, 2, 4) <= 1

Objective

- Given projected TV audience for each game and each week, maximize the total audience

2013 NFL Schedule

GUROBI OPTIMIZATION

end-to-end ANALYTICS

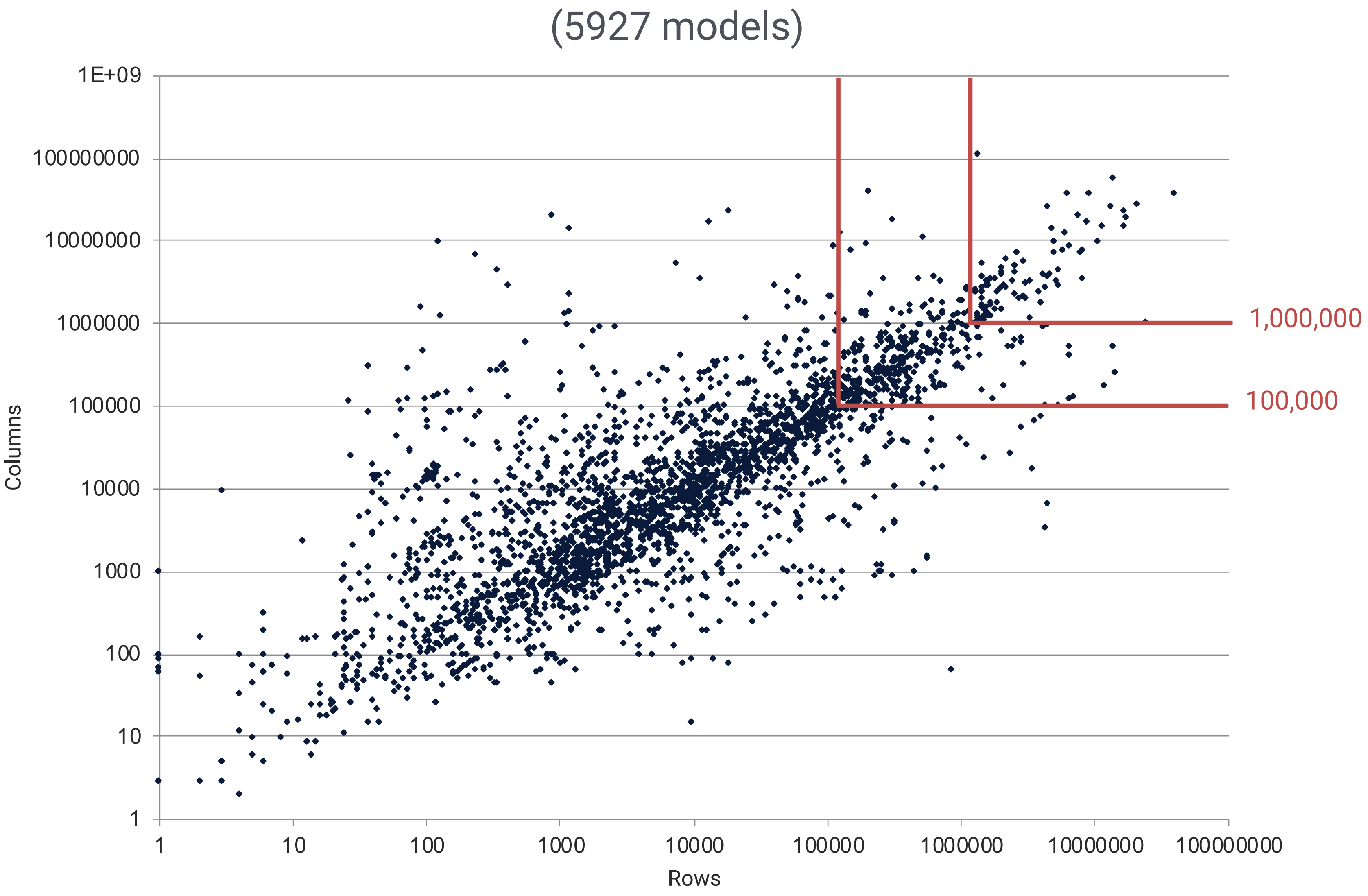
DAL	NYG	PHL	WAS	CHI	DET	GB	MIN	ATL	CAR	NO	TB	ARZ	STL	SF	SEA		BUF	MIA	NE	NYJ	BAL	CIN	CLE	PIT	HOU	IND	JAX	TEN	DEN	KC	OAK	SD
NYG	DAL	WAS	PHI	CIN	MIN	SF	DET	NO	SEA	ATL	NYJ	STL	ARZ	GB	CAR	1	NE	CLE	BUF	TB	DEN	CHI	MIA	TEN	SD	OAK	KC	PIT	BAL	JAC	IND	HOU
KC	DEN	SD	GB	MIN	ARZ	WAS	CHI	STL	BUF	TB	NO	DET	ATL	SEA	SF	2	CAR	IND	NYJ	NE	CLE	PIT	BAL	CIN	TEN	MIA	OAK	HOU	NYG	DAL	JAC	PHI
STL	CAR	KC	DET	PIT	WAS	CIN	CLE	MIA	NYG	ARZ	NE	NO	DAL	IND	JAC	3	NYJ	ATL	TB	BUF	HOU	GB	MIN	CHI	BAL	SF	SEA	SD	OAK	PHI	DEN	TEN
SD	KC	DEN	OAK	DET	CHI		PIT	NE		MIA	ARZ	TB	SF	STL	HOU	4	BAL	NO	ATL	TEN	BUF	CLE	CIN	MIN	SEA	JAC	IND	NYJ	PHI	NYG	WAS	DAL
DEN	PHI	NYG		NO	GB	DET		NYJ	ARZ	CHI		CAR	JAC	HOU	IND	5	CLE	BAL	CIN	ATL	MIA	NE	BUF		SF	SEA	STL	KC	DAL	TEN	SD	OAK
WAS	CHI	TB	DAL	NYG	CLE	BAL	CAR		MIN	NE	PHI	SF	HOU	ARZ	TEN	6	CIN		NO	PIT	GB	BUF	DET	NYJ	STL	SD	DEN	SEA	JAC	OAK	KC	IND
PHI	MIN	DAL	CHI	WAS	CIN	CLE	NYG	TB	STL		ATL	SEA	CAR	TEN	ARZ	7	MIA	BUF	NYJ	NE	PIT	DET	GB	BAL	KC	DEN	SD	SF	IND	HOU		JAC
DET	PHI	NYG	DEN		DAL	MIN	GB	ARZ	TB	BUF	CAR	ATL	SEA	JAC	STL	8	NO	NE	MIA	CIN		NYJ	KC	OAK			SF		WAS	CLE	PIT	
MIN		OAK	SD	GB		CHI	DAL	CAR	ATL	NYJ	SEA		TEN		TB	9	KC	CIN	PIT	NO	CLE	MIA	BAL	NE	IND	HOU		STL		BUF	PHI	WAS
NO	OAK	GB	MIN	DET	CHI	PHI	WAS	SEA	SF	DAL	MIA	HOU	IND	CAR	ATL	10	PIT	TB			CIN	BAL		BUF	ARZ	STL	TEN	JAC	SD		NYG	DEN
	GB	WAS	PHI	BAL	PIT	NYG	SEA	TB	NE	SF	ATL	JAC		NO	MIN	11	NYJ	SD	CAR	BUF	CHI	CLE	CIN	DET	OAK	TEN	ARZ	IND	KC	DEN	HOU	MIA
NYG	DAL		SF	STL	TB	MIN	GB	NO	MIA	ATL	DET	IND	CHI	WAS		12		CAR	DEN	BAL	NYJ		PIT	CLE	JAC	ARZ	HOU	OAK	NE	SD	TEN	KC
OAK	WAS	ARZ	NYG	MIN	GB	DET	CHI	BUF	TB	SEA	CAR	PHI	SF	STL	NO	13	ATL	NYJ	HOU	MIA	PIT	SD	JAC	BAL	NE	TEN	CLE	IND	KC	DEN	DAL	CIN
CHI	SD	DET	KC	DAL	PHI	ATL	BAL	GB	NO	CAR	BUF	STL	ARZ	SEA	SF	14	TB	PIT	CLE	OAK	MIN	IND	NE	MIA	JAC	CIN	HOU	DEN	TEN	WAS	NYJ	NYG
GB	SEA	MIN	ATL	CLE	BAL	DAL	PHI	WAS	NYJ	STL	SF	TEN	NO	TB	NYG	15	JAC	NE	MIA	CAR	DET	PIT	CHI	CIN	IND	HOU	BUF	ARZ	SD	OAK	KC	DEN
WAS	DET	CHI	DAL	PHI	NYG	PIT	CIN	SF	NO	CAR	STL	SEA	TB	ATL	ARZ	16	MIA	BUF	BAL	CLE	NE	MIN	NYJ	GB	DEN	KC	TEN	JAC	HOU	IND	SD	OAK
PHI	WAS	DAL	NYG	GB	MIN	CHI	DET	CAR	ATL	TB	NO	SF	SEA	ARZ	STL	17	NE	NYJ	BUF	MIA	CIN	BAL	PIT	CLE	TEN	JAC	IND	HOU	OAK	SD	DEN	KC

- NBC, CBS, FOX, ESPN, NFLN
- Shaded=home
- Left NFC (E, N, S, W), Right AFC (E, N, S, W)

Partial List of 2013 Rules and Goals

- Teams playing in London home week prior, on BYE week after
- Bills game in Toronto Week 13 against ATL, BAL, CAR, KC, MIA or NYJ
- No team has earliest BYE in consecutive seasons
- No team plays more than two road games against team coming off their BYE
- All teams playing road Thursday games are home the previous week
- Every team plays one Thursday game
- All teams playing home Thursday games have limited travel previous week
- Minimize non-division games during hurricane season (in order to “swap” sites)
- Minimize 3-game road trips (and 3-game home stands for teams with ticket issues)
- Minimize number of division series that end in first half of season
- Minimize number of games that would conflict with MLB postseason
- Maximize the number of late-season division games
- Minimize early-season games 1PM games for teams with weather concerns
- Minimize number of Pacific time zone teams that play at 1PM ET
- CBS/FOX have at least 3 1PM games every week, preferably geographically diverse
- CBS/FOX have at least 5 total games
- Team can play no more the 6 prime time games, and only 3 teams per year can play 6. All other teams can play no more than 5 primetime games (max 4 NBC)

Gurobi MIP Library

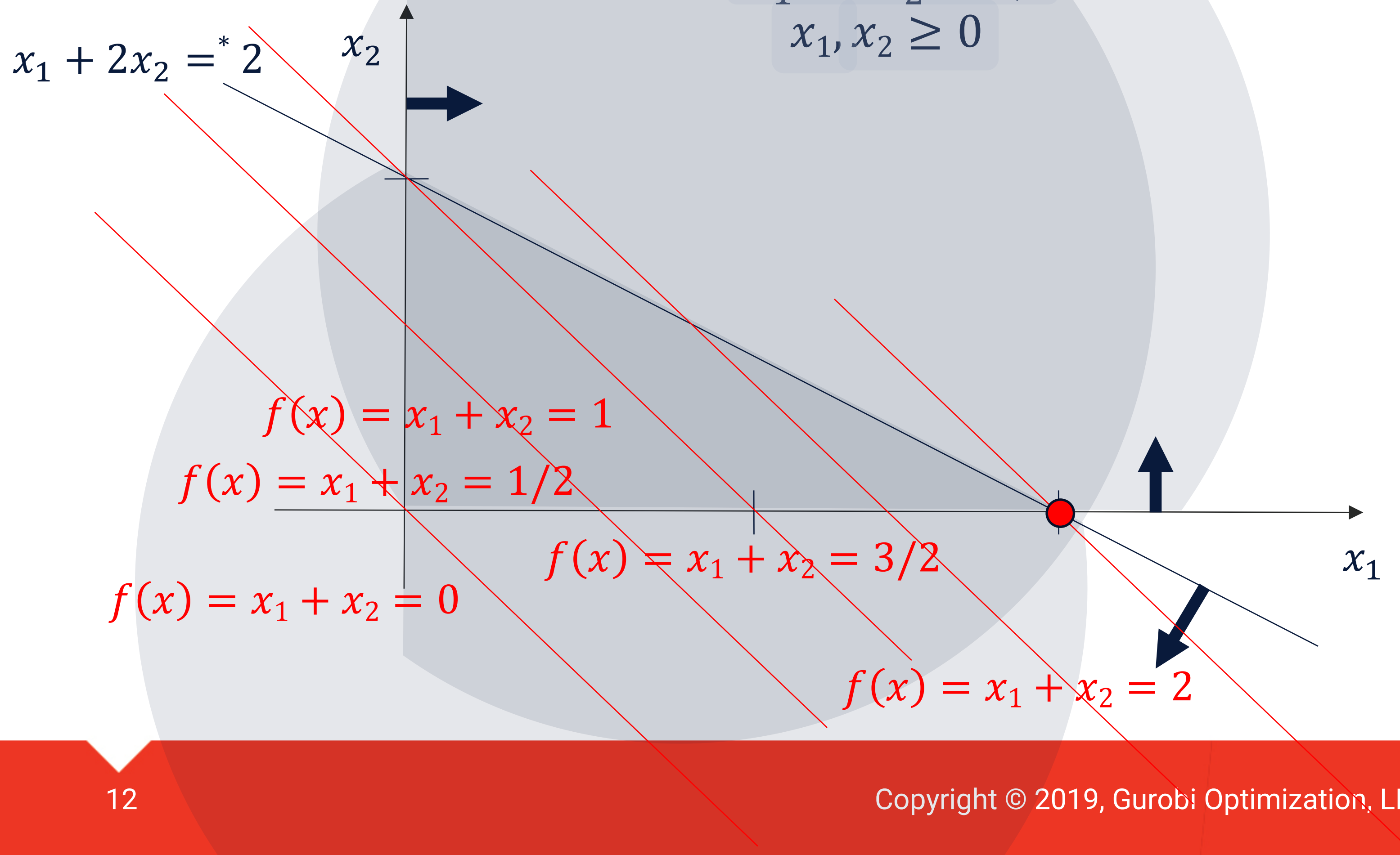


Linear Programming (LP) by Picture

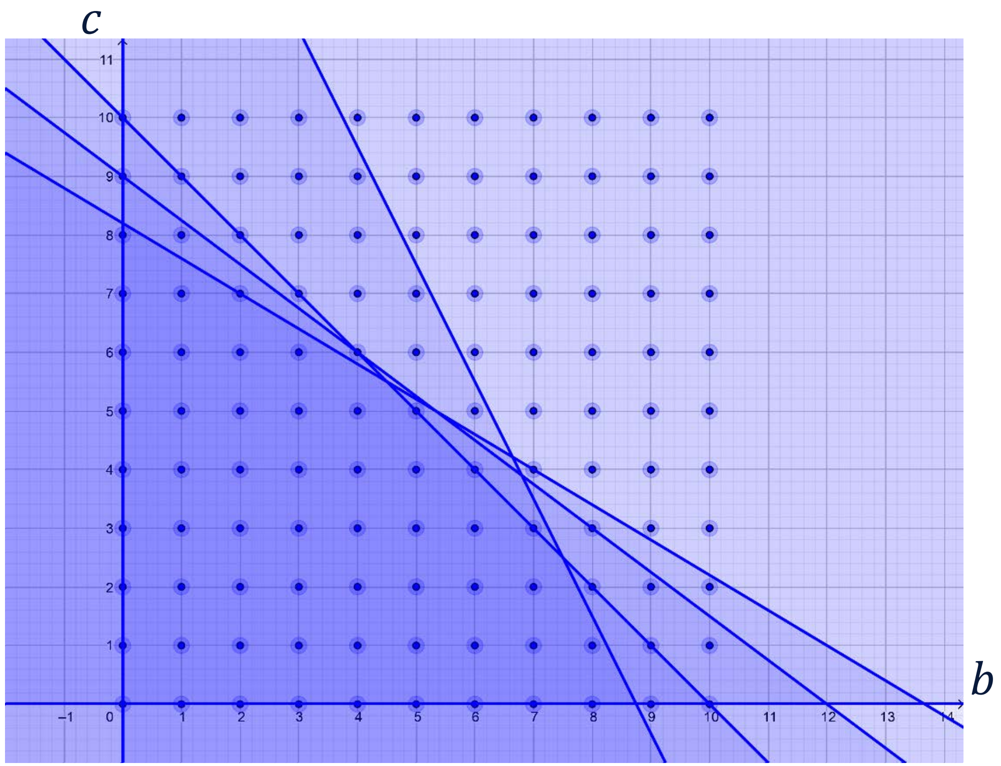
$$\max_{x \in R^2} (x_1 + x_2)$$

subject to

$$x_1 + 2x_2 \leq 2,$$
$$x_1, x_2 \geq 0$$



Mixed-Integer Programming (MIP) by Picture



$$\max_{b,c} \ (10\,b + 11\,c)$$

subject to

$$b + c \leq 10,$$
$$3b + 4c \leq 36,$$
$$3b + 5c \leq 41,$$
$$4b + 2c \leq 35,$$
$$b, c \geq 0$$

Caveat: although may look almost like an LP, algorithmically we are not that good at exploring the integer lattice,

- even a “simple” (0,1)-cube in n-dimensions would have 2^n points! (for instance, in dimension 100 it would have ~1.26×10^{30} points)

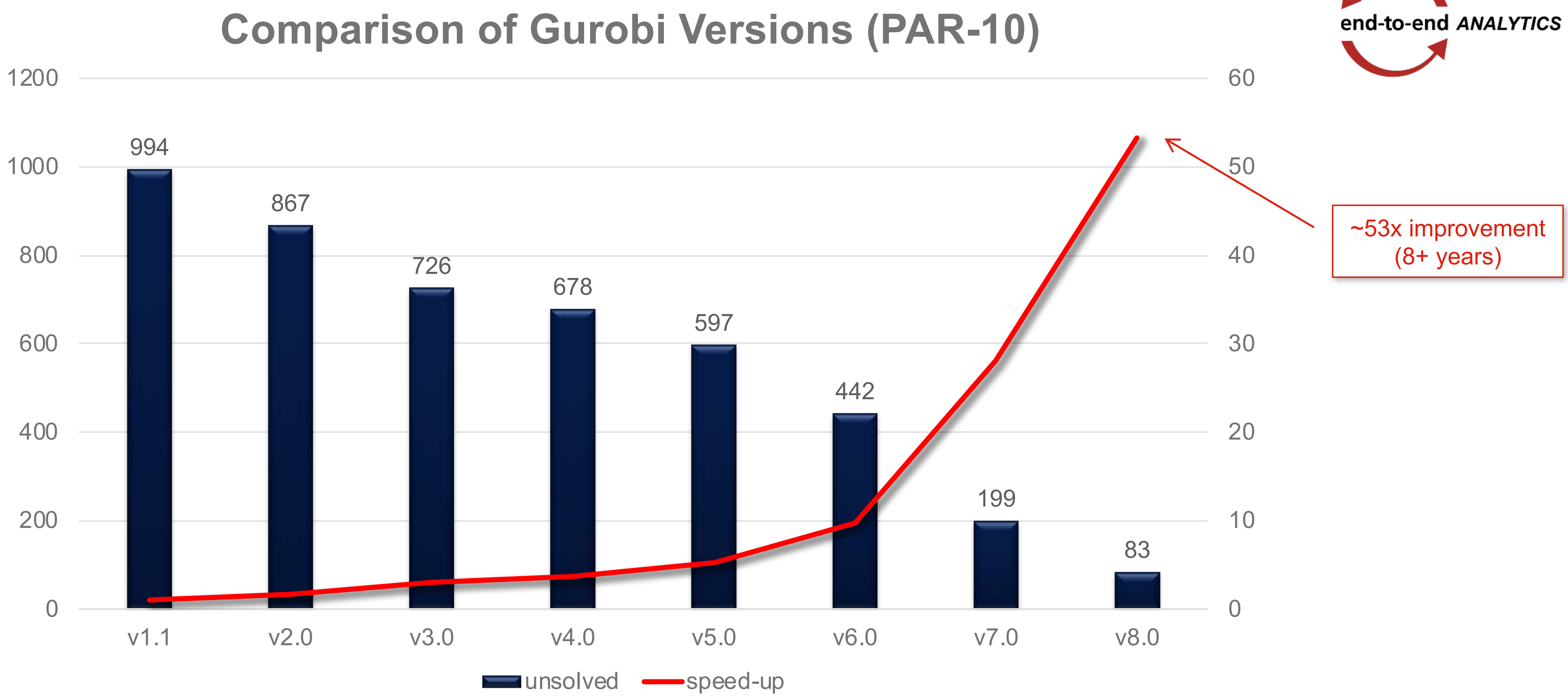
Algorithms for Solving MP Problems

- Simplex (and dual simplex) method for LP
- Interior point algorithms for LP and QP
- Branch and bound for MIP
- Relaxation methods
- Heuristics
- Cutting planes
- Lazy constraints

What Makes MIP Difficult?

- LP vs MIP – polynomial vs NP-complete
- Numerical issues

Importance of Solver Performance



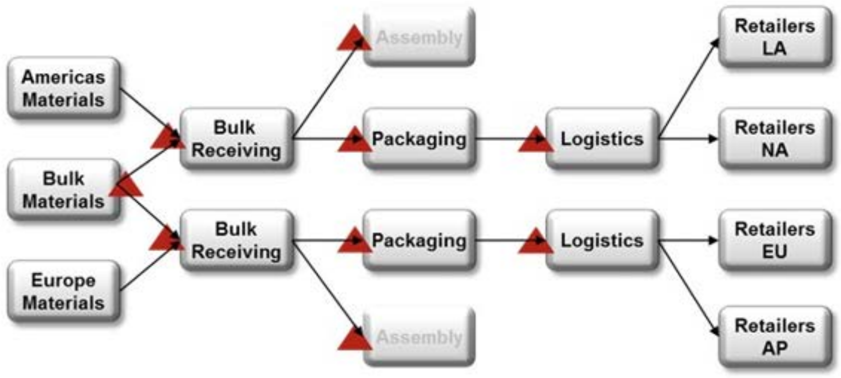
Time limit: 10000 sec.
Intel Xeon CPU E3-1240 v3 @ 3.40GHz
4 cores, 8 hyper-threads
32 GB RAM

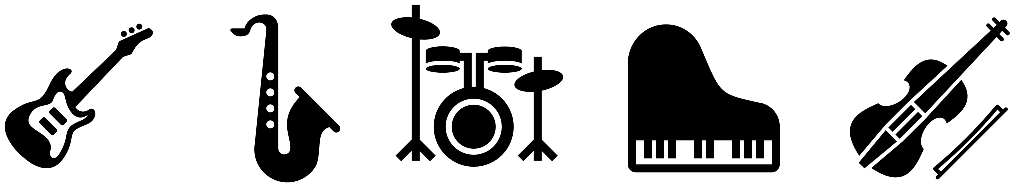
Test set has 5656 models:
- 410 discarded due to inconsistent answers
- 1493 discarded that none of the versions can solve
- speed-up measured on >100s bracket: 2012 models

Identifying Optimization Problems Within Your Organization

Dr. Russell Halper

When Does Optimality Matter?





Designing a Distribution Network

Challenge: Where do I put my assets and how do I move products between them?

Decision: A selection of sites to build assets, type of asset, and a flow of products to/from each

Cost of a Single Sub-Optimal Decision: Often times, tens of millions of dollars… or more!

Music Recommendation

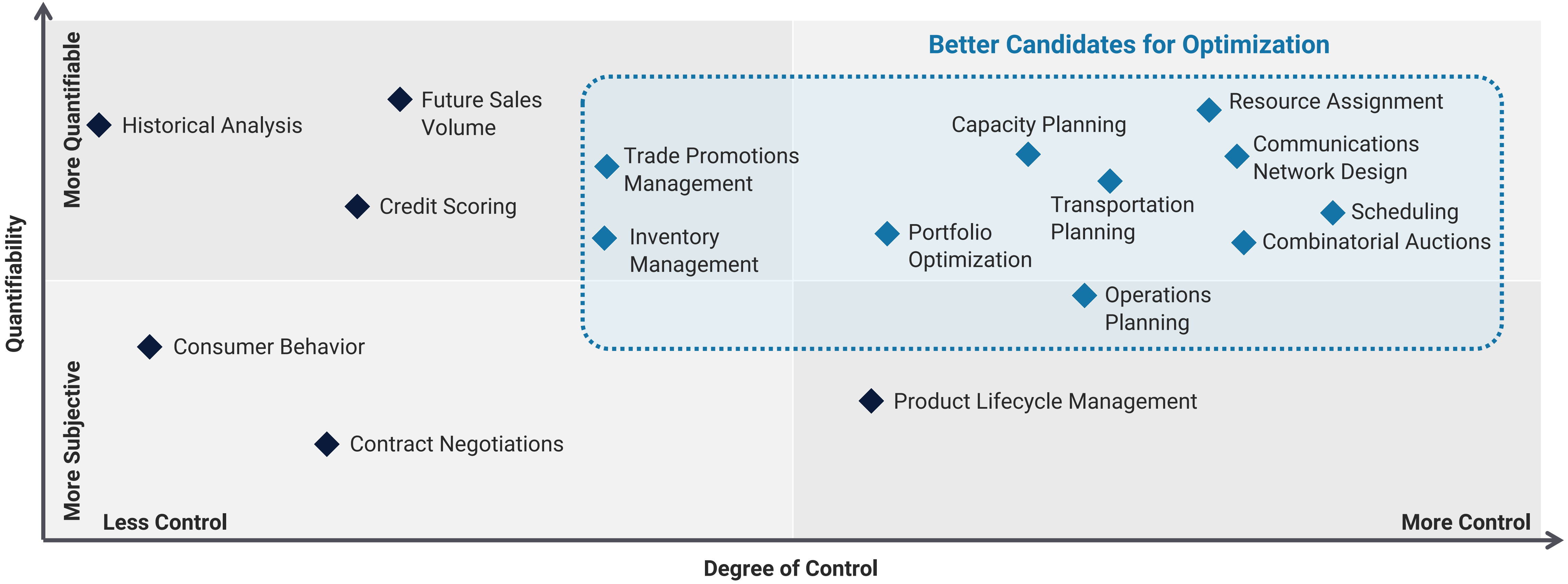
Challenge: What songs would a customer enjoy?

Decision: A selection of one or more songs to play to a customer

Cost of a Single Sub-Optimal Decision: Negligible

Identifying Opportunities for Optimization

Optimization can often be used when the business has a high degree of control over a reasonably quantifiable business problem.



How MIP Complements Machine Learning

Dr. Russell Halper

Many ML Techniques Use *Some* Type of Optimization

GUROBI OPTIMIZATION

end-to-end *ANALYTICS*

PREDICTING A VALUE				SEGMENTING	
Predict a numerical value Is the predicted value linear in the data?		**Predict a categorical value** Is the predicted value linear in the data?		**Group together like data** Is the number of groups known?	
Linear Methods	**Nonlinear Methods**	**Linear Methods**	**Nonlinear Methods**	**Known**	**Unknown**
Linear Regression Prior knowledge and/or small data sets: **Bayesian Linear Regression** Feature reduction: **LASSO, Ridge**	Interpretability: **Regression Trees** Accuracy: **Random Forests** Large data sets: **Neural Network Regression**	Probabilistic Approach: **Logistic Regression** Explicitly create boundaries: **Support Vector Machines**	Interpretability: **Classification Trees** Accuracy: **Random Forests** Large data sets: **Neural Network Classification**	Each data point in its own cluster: **K-Means, K-Median** Overlapping clusters: **EM Clustering**	Always gives the same result: **Hierarchical clustering** **Affinity Propagation**

Many ML Techniques Use *Some* Type of Optimization

PREDICTING A VALUE				SEGMENTING	
Predict a numerical value Is the predicted value linear in the data?		Predict a categorical value Is the predicted value linear in the data?		Group together like data Is the number of groups known?	
Linear Methods	Nonlinear Methods	Linear Methods	Nonlinear Methods	Known	Unknown
Linear Regression Prior knowledge and/or small data sets: **Bayesian Linear Regression** Feature reduction: **LASSO, Ridge**	Interpretability: **Regression Trees** Accuracy: **Random Forests** Large data sets: **Neural Network Regression**	Probabilistic Approach: **Logistic Regression** Explicitly create boundaries: **Support Vector Machines**	Interpretability: **Classification Trees** Accuracy: **Random Forests** Large data sets: **Neural Network Classification**	Each data point in its own cluster: **K-Means, K-Median** Overlapping clusters: **EM Clustering**	Always gives the same result: **Hierarchical clustering** **Affinity Propagation**

Regression

Minimize the sum of squared errors when fitting the model

Neural Networks

Stochastic Gradient Descent is used to minimize the average error over the training data set

K-Median/K-Means

Assign a data point to a cluster to minimize some aggregate distance metric

K-Median Clustering as a MIP

From a set ***P*** of data points, select a subset of ***k*** points (i.e., “centers”) that minimizes the distance from each point in ***P*** to its closest center

Index Sets

P = the set of points in the model

Data

$Dist_{ij}$ = Distance between points i and j

Variables

$$y_j = \begin{cases} 1 \text{ if point } j \text{ is a Center} \\ 0 \text{ otherwise} \end{cases}$$

$$x_{ij} = \begin{cases} 1 \text{ if point i is assigned to center } j \\ 0 \text{ otherwise} \end{cases}$$

Minimize: $\sum_{i,j \in P} Dist_{ij}\, x_{ij}$ — Minimize the Distance

Subject To:

$\sum_{j \in P} y_j = k$ — Select *k* centers

$x_{ij} \le y_j$ for all $i, j \in P$ — Only assign points to a selected center

$\sum_{j \in P} x_{ij} = 1$ for all $i \in P$ — Each point must be assigned to exactly one center

$x_{ij}, y_j \in \{0,1\}$ for all $i, j \in P$ — Variables are either 0 or 1

Predictive Models Often Require Actions to Be Taken Before Value is Realized….

GUROBI OPTIMIZATION

end-to-end *ANALYTICS*

Descriptive	Predictive	Prescriptive
What happened and why?	What will happen and why?	What should the organization do?

Predictive ML

Many excellent options!

Interpretable:

- Lasso
- Random Forests
- Classification Trees

Non-Intrepretable:

- Neural Networks
- Reservoir Computing

Prescriptive

- **Mathematical Programming (Optimization):** Provably optimal results. High degree of control over problem.
- **AI/Neural Networks:** Evolving technology and useful for many problems. Black box.
- **Simulation:** Useful for non-convex problems with multiple local minima or highly-linear problems
- **Heuristics:** Obtainable but can be inflexible when new constraints added to a business problem

Provably optimal solution (Mathematical Programming)

Generally not provably optimal solutions (AI/Neural Networks, Simulation, Heuristics)

Should I Use MIP, AI, or Another Approach?

GUROBI OPTIMIZATION

end-to-end ANALYTICS

- How important is optimality?
  - Is the cost of sub-optimal solutions high?
  - Most optimization models are NP-Complete problems. Polynomial runtime algorithms cannot find an optimal solution.
- Does data exist to train a model? Is it a new situation? Will the business problem change?
- What type of scenarios analysis needs to be done? What type of sensitivity analysis needs to be done?
- Do I need more than one solution?
- How big is the problem?

Optimization and Machine Learning Working Together

GUROBI OPTIMIZATION

end-to-end *ANALYTICS*

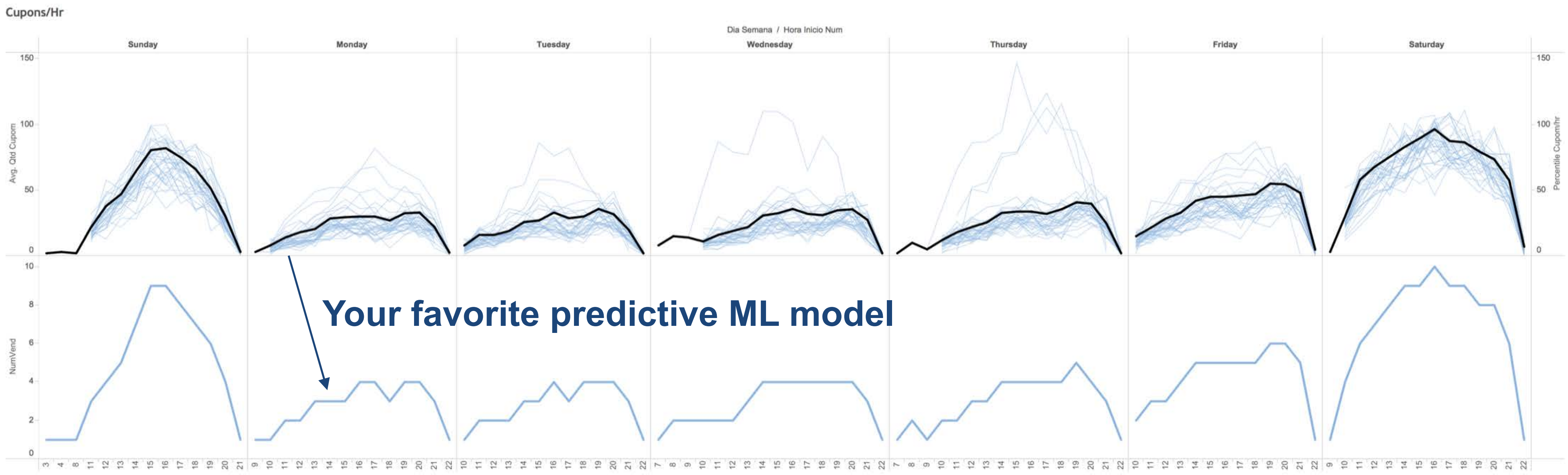
- **Retail chain w/ ~200 stores**
- **+1400 Employees**
- **Demand varies throughout the day**
- **Complicated local labor laws**
- **Problem statement:** How much staff should we hire for each store?

	1-Monday	2-Tuesday	3-Wednesday	4-Thursday	5-Friday	6-Saturday	7-Sunday	1-Monday	2-Tuesday	3-Wednesday	4-Thursday	5-Friday	6-Saturday	7-Sunday	1-Monday	2-Tuesday	3-Wednesday	4-Thursday	5-Friday	6-Saturday	7-Sunday
1048.1.1		8	8	8		10	8		8	8		8	8	8	8	8	8			8	
1048.1.13				8	8	10	8		8				8					8	8	10	8
1048.1.18						10	8		8	8			10			8	8	8		8	8
1048.1.2			8	8	8	8	8			8	8	8	10	8	8		8	8	8	8	
1048.1.21	8	8			8	10		8		8		8	8	8		8	8	8		8	8
1048.1.26	8			8	8	8	8					8	10	8					8	8	
1048.1.29	8	8			8	8	8	8	8		8	8	8			8		8	8	10	8
1048.1.3					8	8		8	8			8	10	8				8	8	8	8
1048.1.31	8		8	8	8	8	8	8		8	8	8	10		8		8	8	8	8	8
1048.1.35		8	8	8		8	8			8	8	8	10	8	8			8	8	8	
1048.1.38	8	8			8	8						8	10	8		8		8		10	8
1048.1.39	8	8		8		8	8		8		8	8	10		8	8		8	8	8	8
1048.1.4	8	8	8		8	10	8	8	8		8	8	8	8	8		8	8	8	10	
1048.1.42			8	8	8	8	8		8	8	8		8		8	8			8	10	8
1048.1.44						10		8	8		8	8	8	8					8	8	8
1048.1.5		8	8		8	10							8	8			8		8	10	8
1048.1.6	8		8	8	8	8	8	8		8	8		10	8	8	8	8		8	10	
1048.1.7		8	8	8	8	10		8		8	8	8	8	8		8	8	8	8	10	8

Developing a Predictive Model

So how does one predict how much labor in needed?

- Register sales provided a key piece of data that help to understand the level of staffing over time.
- Other key demand drivers: Day of week, Time of Day, Holidays, Seasonality, Weather, ….
- What does the business need out of a predictive model:



What Does the Optimization Model need to Consider?

Determine how many employees to hire (minimize)

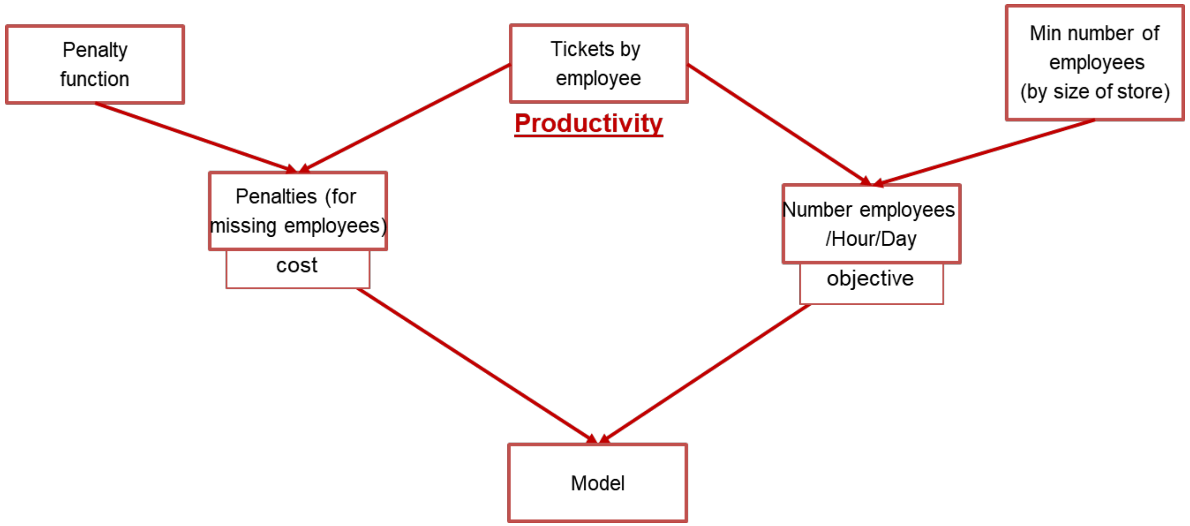
Schedule (what days, what hours, when to break)

Employees are of 2 types:

- Regular employees
  - Fixed monthly salary
  - Can be scheduled at most 6 consecutive days
  - Can be scheduled at most 2 consecutive Sundays
  - Same starting hour during weekdays
  - Fixed number of hours per day (9 hours)
  - Overtime is available (at additional cost)
  - Min and max number of hours per week
  - During a day, cannot work more than 6 hours without a 1-hour break\
- Part-time employees
  - Salary determined by assigned hours
  - Less restrictions of consecutive days or Sundays
  - Flexible starting hours
  - No overtime is available

Model Penalties

- The model penalizes if there are fewer employees are assigned than the requirement (by hour, day)
- It is assumed that fewer employees will result in loss of sales



$$Total\ Cost = Salaries + Penalties + Overtime\ costs$$

Resources

- **Access www.gurobi.com and browse through Resources on our website**
  - This includes code examples, webinar recordings, documentation and information on how to get started with optimization.
- **Get a free 30-day trial of Gurobi**
  - www.gurobi.com/eval
- **Join us in next week's webinar (May 21 & 22):**
  - Optimization Application Demos - This presentation will walk through two live demos implemented by the Gurobi team and deployed on Amazon Web Services using Docker and Gurobi Instant Cloud.
  - Register now: https://bit.ly/2LshNki
- **Need help leveraging Optimization for your business?**
  - Contact russell@e2eanalytics.com.