

A/B Testing and Beyond

Designed Experiments for Data Scientists



Continuing Education Certificate offered by:



UNIVERSITY OF
SAN FRANCISCO



Why, you ask?



Find your job

Menlo Park



Data & Analytics

Research Scientist, Experimental Design and Causal Inference

Menlo Park, CA

Apply

Please limit to 3 applications.



Find your job

Menlo Park

Data & Analytics

Research Scientist, Experimental Design and Causal Inference

Menlo Park, CA

Apply

NETFLIX

Please apply

Data, Analytics, and Algorithms

Science and Algorithms

Los G

ALL JOB OPENINGS

Senior Data Scientist - Streaming Experimentation and Modeling

Los Gatos, California



Find your job

Menlo Park

Data & Analytics

Research Scientist, Experimental Design and Causal Inference

Menlo Park, CA

Apply

NETFLIX

Please apply

Data, Analytics, and Algorithms

Science and Algorithms

Los Gatos



Our Company

Stories

Global

Start Your Search

Find Jobs by Category

Find Jobs by Location

Engineering at eBay

Students & Recent Graduates

Meet Our People

Our Benefits

How We Recruit

Contact Careers

Job Search Results

Find your dream job

Locations

Experimentation Science

San Jose, California

Apply

Experimentation and

Find your job

Data & Analytics

Research Scientist, Experimental I

Menlo Park, CA

Apply



Start Your Search

Find Jobs by Category

Find Jobs by Location

Engineering at eBay

Students & Recent Graduates

Meet Our People

Our Benefits

How We Recruit

Contact Careers

Find your dream job

Experimentation Science

San Jose, California



Microsoft Careers

Working at Microsoft ▾ Students and graduates ▾

Principal Data Scientist

The Core Data Science team of the Windows and Devices Group (WDG) is responsible for a broad range of data analysis, machine learning and engineering insights in the Windows ecosystem. We use comprehensive device telemetry, service information and external data sources to improve the quality of experiences on our devices and services for PCs, Surfaces, tablets, Xbox, mobile devices, HoloLens, browsers, apps, and services. We pioneer new methods, data integrations, instrumentation and operational processes to lead the transformation of WDG into a truly data-driven engineering organization.

Our team was founded at the beginning of the Windows 10 project to lead the way for the engineering team to adopt strong data and analytic skills as daily competencies for the product team. We have performed hundreds of analyses with concrete results for customers since then. We are a large, multi-disciplinary team of experts working on one of the largest and most varied data streams in the world with major impact to Windows business. The team collaborates with the broader community of data and applied scientists across Microsoft and publish our work across the company weekly.

We are looking for a skilled and experienced Principal Data Scientist with a deep background in systems engineering, experimental design, statistics, simulation or a related field to focus on driving the use of experimental techniques, including optimal experimental design. This role is key to help with the transformation to deliver on

LIP

Los C

and

What is this all about?

- *Designed experiments* are a foundational tenet of data science
- Machine learning is cool, but...
- With experiments we can identify and quantify cause-and-effect relationships



What is an experiment?

- An experiment is a planned investigation in which the influence of one or more factors on a response variable of interest is quantified
- Here data collection is *purposeful* as opposed to *transactional*
- It facilitates causal inference



What designs will we consider?

- Experiments with 2 variants
- Experiments with n variants
- Multi-factor multi-level experiments
- Response surface methodology
- Multi-armed bandit experiments



Intriguing, tell me more...

- Carefully designed experiments can answer a host of important questions
- Experimentation can be used for
 - Screening
 - Optimization



Screening Experiments?

- Given a large number of factors, which ones are *active*? i.e., which ones have a large impact on the response?
- **Example:** In an email advertising campaign interest lies in determining which factors most influence conversion.
 - Ad type
 - Discount offer
 - Discount duration
 - Time of day
 - Day of week
 - Email cadence



Optimization Experiments?

- Given a handful of highly influential factors, what levels should they be set at to maximize or minimize a response?
- **Example:** Your favorite ride-sharing app performs a screening experiment and finds discount amount and duration are the most influential factors on conversion. What “levels” should these factors be set at to maximize conversion?
 - Discount amount: 10%, 30%, 50%
 - Discount duration: 1, 3, 5 days



The Ride-Sharing Example

Goal: maximize effectiveness of incentive program (i.e., maximize ride bookings during experimental period)

Response: Booking rate

Factors: Discount amount (10%, 30%, 50%)
Discount duration (1, 3, 5 days)



The Ride-Sharing Example

Experimental Setup:

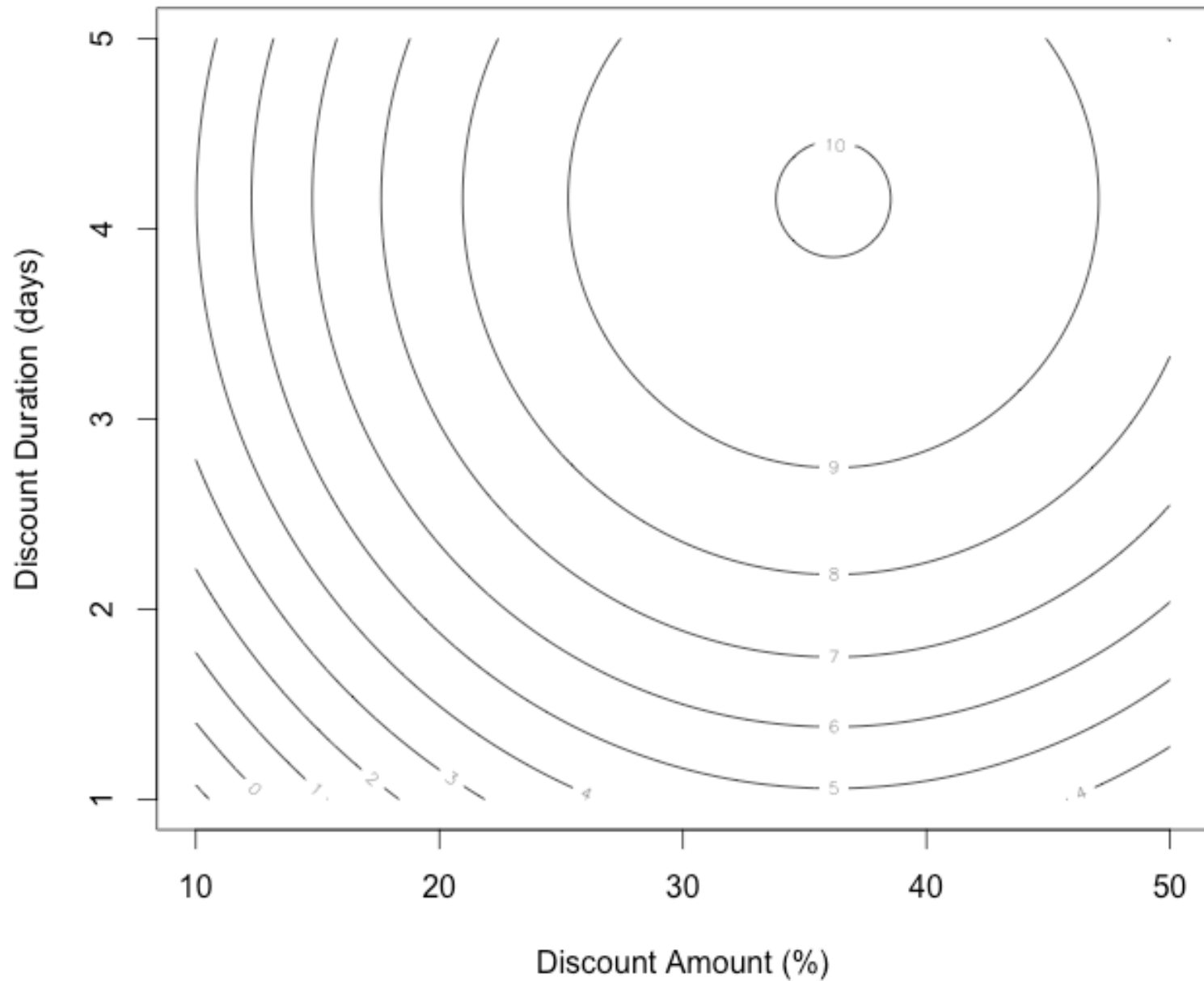
- 5 conditions

	10% Discount	50% Discount
1-day Duration	#1	#2
5-day Duration	#3	#4

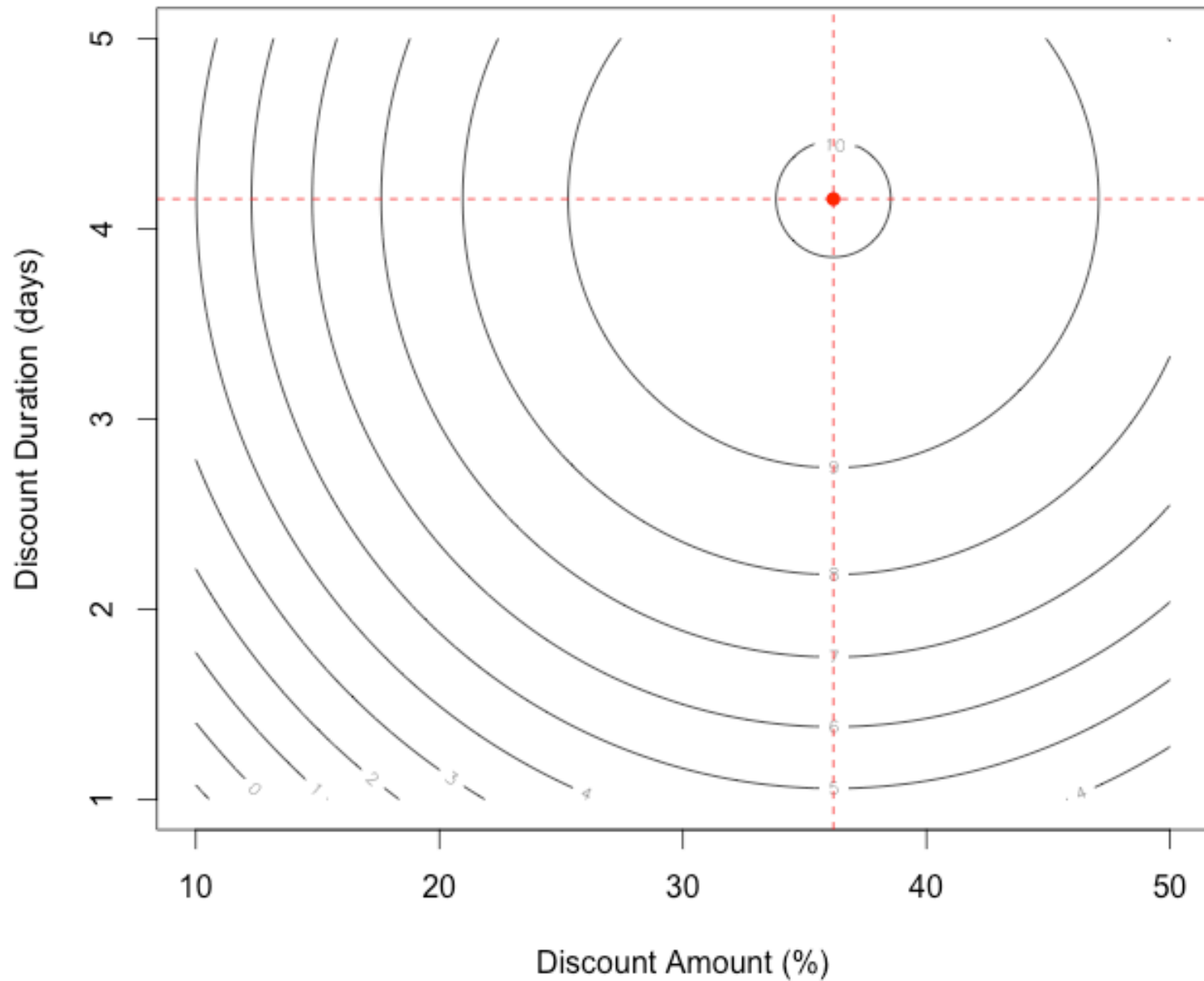
#5 is positioned in the center, overlapping the intersection of the 10% Discount and 50% Discount columns and the 1-day Duration and 5-day Duration rows.

- Randomly assign 1500 users to each of the 5 conditions
- Measure the proportion of the 1500 users that book during this experimental period
- Fit a response surface model to the data

Booking Rate by Discount Amount and Duration



Booking Rate by Discount Amount and Duration



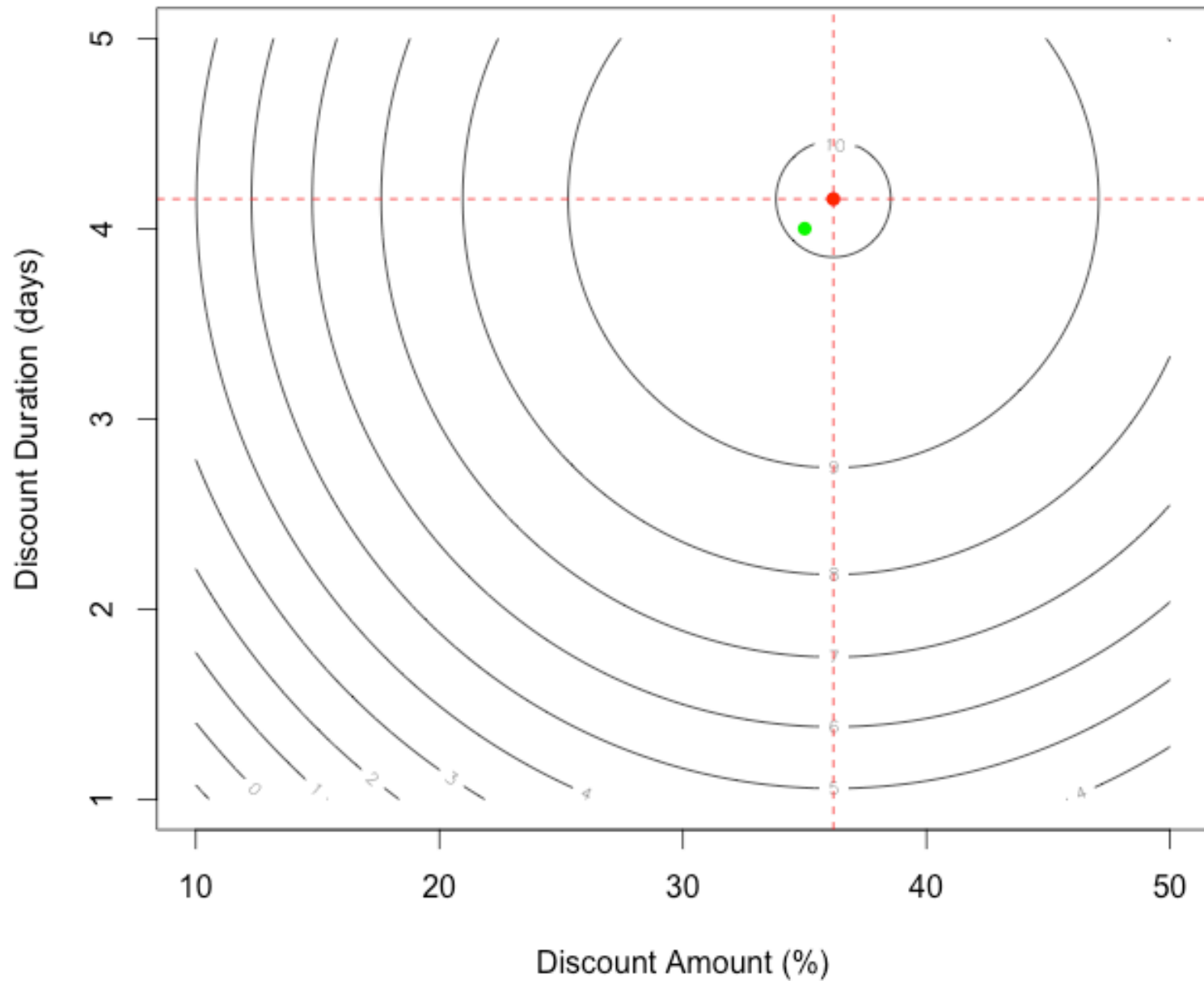
The Ride-Sharing Example

Optimal Discount Settings:

- Discount Amount:
 - 36.17863%
- Discount Duration:
 - 4.155968 days
- Resultant Booking Rate:
 - 10.04902%



Booking Rate by Discount Amount and Duration



The Ride-Sharing Example

What about something more practically feasible?

- Discount Amount:
 - 35%
- Discount Duration:
 - 4 days
- Resultant Booking Rate?
 - 10.02412 %



Administrative Details

Dates:

- Wednesdays 6:00-9:00pm
- September 6 – October 18, 2017

Location:

- 101 Howard Street, San Francisco, CA

Cost:

- \$1590



Administrative Details

Lecture Style:

- 2 hour lecture + discussion
- 1 hour in-class lab

After Lecture Work:

- Assigned exercises
- Completion of exercises required for successful completion of certificate



Enroll Now!

<https://www.usfca.edu/data-institute/certificates/testing-designed-experiments-for-data-scientists>

For questions, email me:

ntstevens@usfca.edu

