Swim, Bike, Run
Keep Data Science Simple
Session Track: Analytics & AI

John Young
Head of Data Science & Machine Learning
Email: jyoung@hitachisolutions.com
#HSCCATLANTA19
This is not a recipe for simplifying data science, it is *not* simple.

This will inspire you to keep your connection to the basics.

First, a story...
Where should I start?

ORDER IS IMPORTANT

Right priorities

Right structure

Right technology
Customer Check *Priorities*

Swim 2.4miles + Bike 112miles + Run 26.2miles
Develop Use Case

What:
- 1st Ironman
- 1st Triathlon
- 1st Marathon
- 1st Open water swim (not in a pool)
- 2000 start in the water

Measure of success:

Finish under 17 hrs
Which opportunity?

Not all machine learning or data products are the same. Start with these two guidelines.

1. Can a human do it at all?

2. Can many humans do it reliably?

• If Yes=2 then let’s consider it...
Some data for you...

Problem space:

Death Rate: **1.5 per 100,000** (2x Marathon rate, 65%)

Ironman Finishers: **700,000 in 7.3B** (0.01% est)

Annual Ironman events worldwide: **36+**

Originated: **1978 Hawaii**

Avg Income: **$126,000 /yr** (40-49yr highest group)

...
Customer Check *Structure*
What do we have?

Structuring an ML project is not like appDev, it is product development.

1. Data
2. Talent
3. Leadership
More data for you...

Who: John
Age: 33
Gender: Male
Longest bike: 200km
Longest run: 12mi
Fastest 5k: 17:49min
Longest swim: 100yd
More data for you…

Observations:
Triathlon popularity has leveled off
Differences in finishing between venues is small
Normal distribution of finishers between 9 and 17hrs

Guidance:
Pick an event that is convenient – Wisconsin!
Maintain strong bike endurance
More data for you...

Observations:
Strong correlation between run time and finish time
Finish times are consistent over last 10 years
Going too fast on the bike can hurt your run!

Guidance:
Relax during the swim
Long training runs should be completed while fatigued
Conserve energy during bike
Customer Check Technology
Inventory

What:
- Food
- Goggles
- Wetsuit
- Running shoes
- Bike & Gear

Constraints:
- Limited $, training time
Data Science Inventory

- Data Storage w/Data Model (client specific)
- Connected Sensors
- Event Processing Resources
- "Defined Intelligence & Machine Learning"
- Visualizations live, real-time Inventory custom alerting notifications
- Event Hubs IoT Hub
- Role Based Access Control
- Azure Active Directory
- Excel file input & internal data (Finance SAP)
- ERP Customer/other data

Data Lake
- SQL Data Warehouse

Analytics
- Shim code, if needed
- AzureML
- R or Python, Jupyter Notebooks
- Storage blob
- DocumentDB or different Storage blob
- PowerBI Dashboard Application Insights

Connected Devices
- Intelligent Devices
- Sensor reference data

Data Factory
- Stream Analytics

Remote Monitoring
- Operationalized Management Portal

Auto-Replenishment
- Predictive Maintenance
- Operationalized Management Portal
Step One
Swimming
- Know how to swim
- Get in the water
- Have an exit strategy, you can die underwater.

Step Two
Biking
- Get a bike that works
- Practice ½ distance
- Save something for the Run

Step Three
Running
- You can walk
- Keep moving
- You will want to do it again
It doesn’t take a gifted athlete to do an Ironman, it takes a dedicated one.

- Ted Shue, colleague and friend
Step One
Get Data
- Sample tables, docs, images
- SQL, Scala, Dplyr, Pandas
- Limit your time

Step Two
Explore Data
- Visualize
- Clustering
- Correlation

Step Three
Discuss Findings
- Engage your business frequently
- Accept what data tells you
- Present where you want to go next
It doesn’t take a gifted data scientist to build a product, it takes a dedicated one.

- Someone somewhere
As part of the Chicago Microsoft Technology Center partnership, the Capax team partnered with MTC management and created a visual display that is currently staged in the MTC.

- Customer video testimonial
- PowerBI dashboards
- Demonstrates the power and ease of getting started with an IoT project on the Microsoft platform

Analyzing the telemetry data that is gathered from the activity of the cyclist provides insights into the overall health of the athlete in addition to predictive analytic insights that captures:

- Body temperature monitoring
- Heart rate
- Movement (sway) for form and energy monitoring
- Geo location
- Power (in watts)
- Linear Acceleration
Thank you!
Jyoung@hitachisolutions.com

Finish time: 13hr 15min