ANALYSIS, VISUALIZATION AND INTERPRETATION OF BIOMETRIC DATA FROM MOBILE SMART DEVICES

Hadley Myers , JMP

PhUSE SDE, Utrecht, 7 February 2018

INTRODUCTION

ANALYSIS, VISUALIZATION AND INTERPRETATION OF BIOMETRIC DATA FROM MOBILE SMART DEVICES

- Process:

1. Smart device collects data.
2. Access corresponding data.
3. Compile data.
4. Clean data.
5. Exploratory visualization
6. Predictive modeling.
7. Engagement of stakeholders.
8. Scale-up and automation.

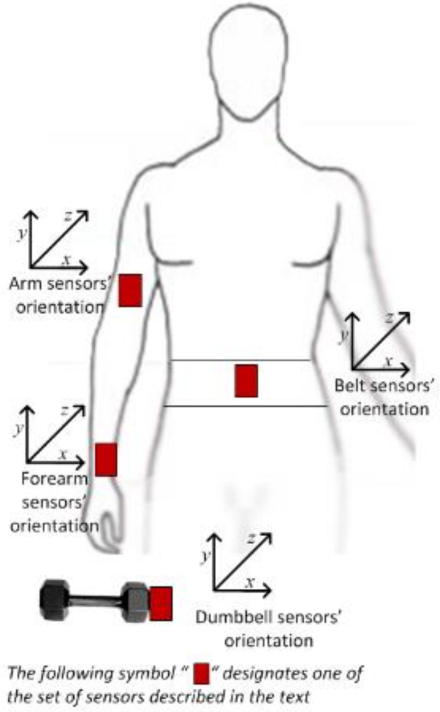
Statistical Discovery.™ From SAS.

TRACKING WORKOUT METRICS

- Cardio metrics:
  - Time
  - Distance
  - Steps
  - Routes (GPS)
  - Type of workout (steady-state, interval, etc).
- Resistance training metrics:
  - Lifts or movements
  - Weight used
  - Max weight lifted (can be risky!)
- Other Data:
  - Date/Time
  - Weather
  - Scores
  - Meals

EXAMPLE

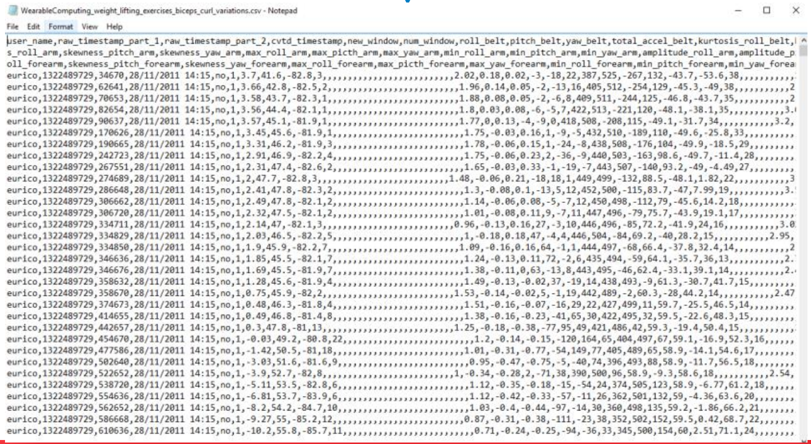
- Weight Lifting Exercises Dataset
  - Data on Unilateral Dumbbell Bicep Curls
    - Participant asked to perform bicep curls
    - Curls watched by expert and judged:
    - Good or Bad
  - Goal is to use metrics from sensors to identify type of lift.

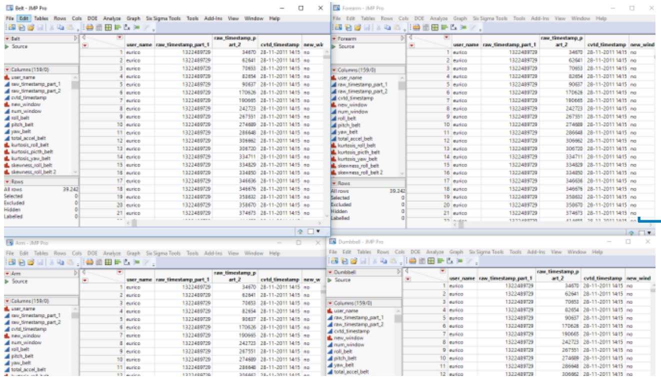


Source: Velloso, E.; Bulling, A.; Gellersen, H.; Ugulino, W.; Fuks, H. **Qualitative Activity Recognition of Weight Lifting Exercises**. Proceedings of 4th International Conference in Cooperation with SIGCHI (Augmented Human '13) . Stuttgart, Germany: ACM SIGCHI, 2013.

DATA COLLECTION

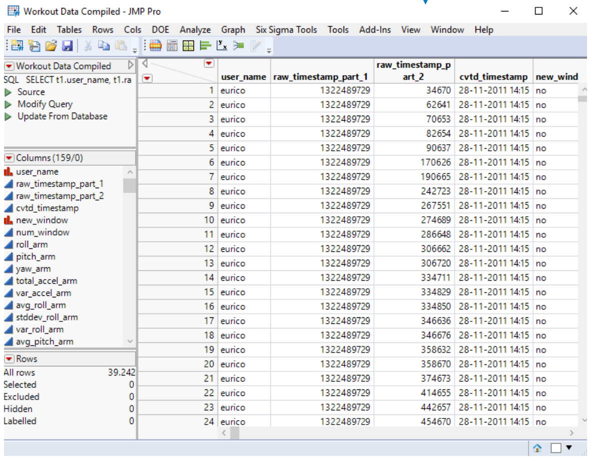
Data imported from wearables in CSV format.





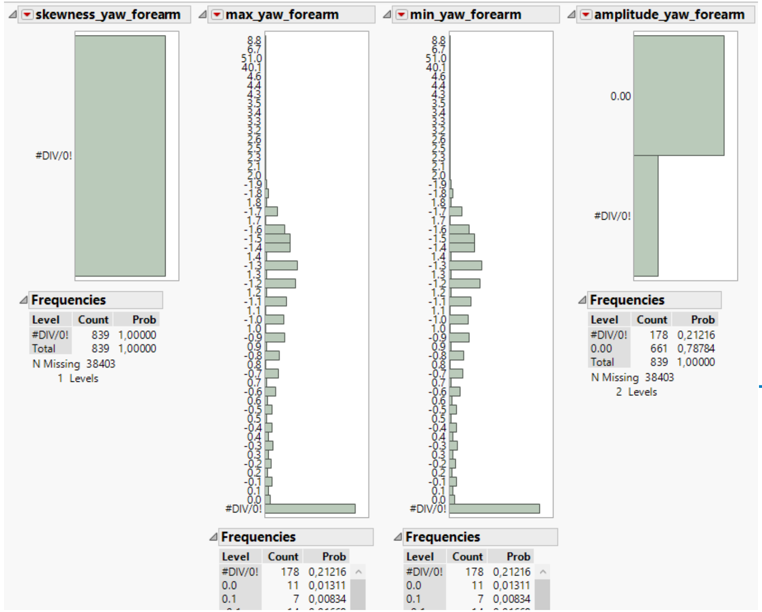
Data from separate wearables merged to create combined dataset.

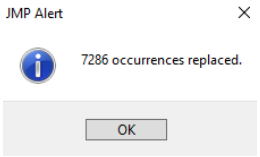
CSV converted to analysis platform of choice

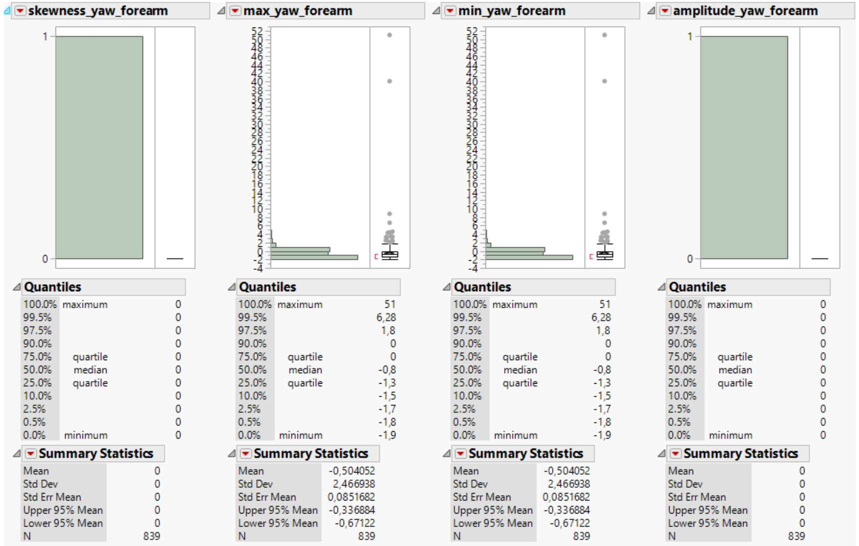


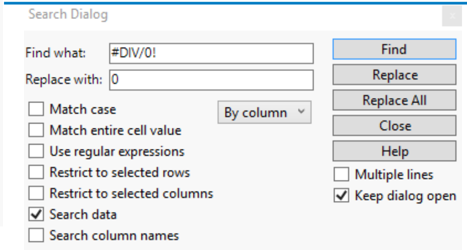
THE POWER TO KNOW.

DATA CLEANING



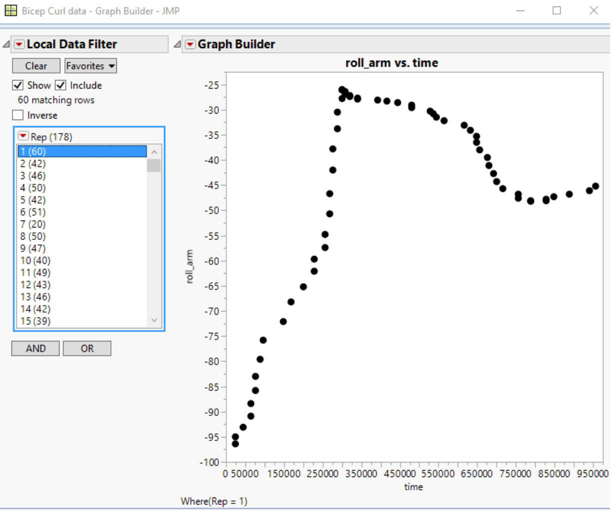


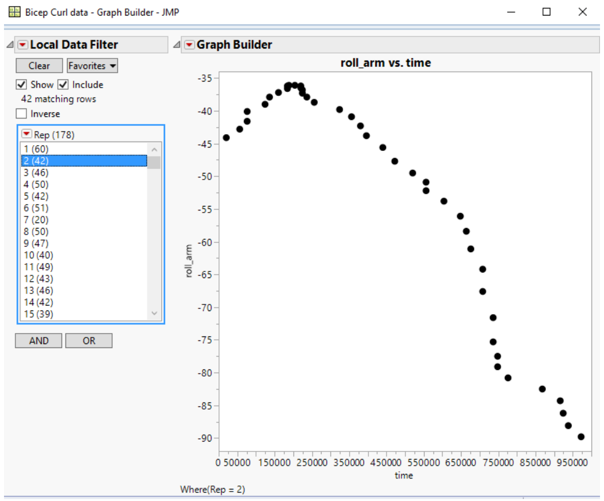


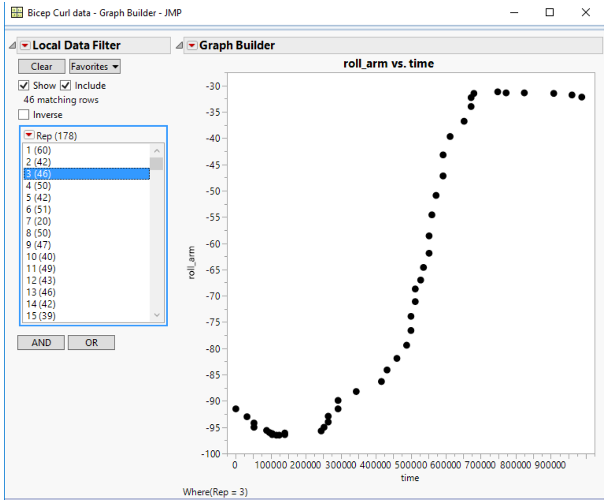


- Erroneous symbol removed.
- Columns converted from nominal to numeric.

EXPLORATORY DATA VISUALIZATION

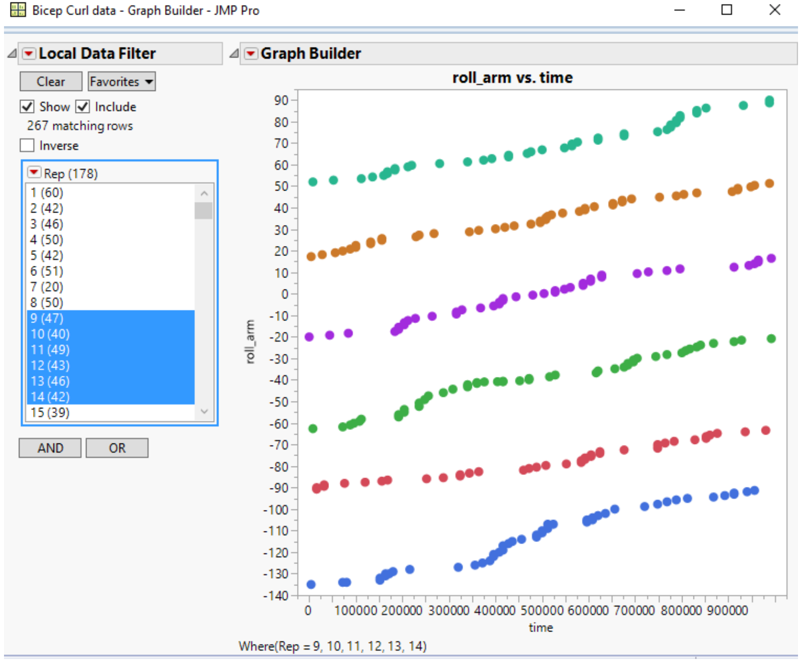


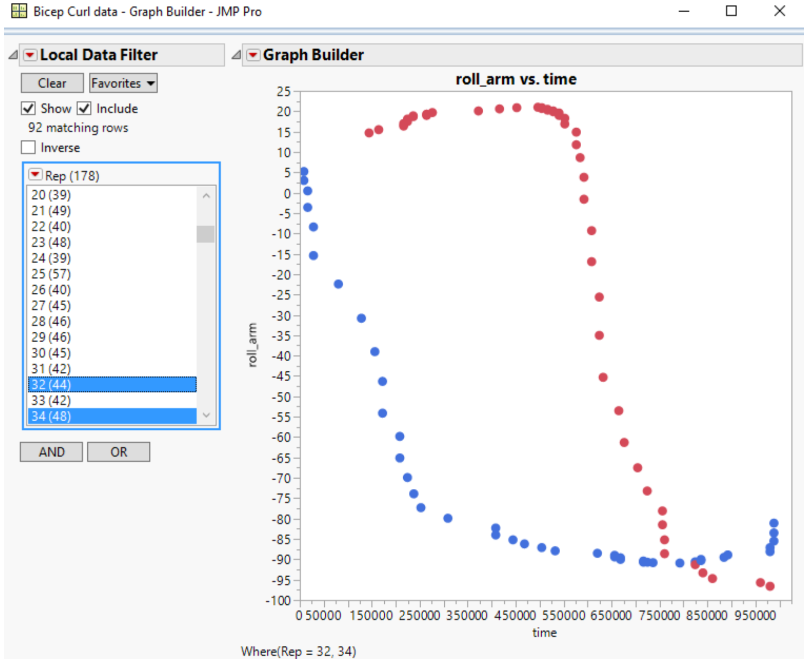




Use of local data filter reveals structure by time

EXPLORATORY DATA VISUALIZATION





Looking at only “Good” reps, we perhaps notice a trend (positive slope).

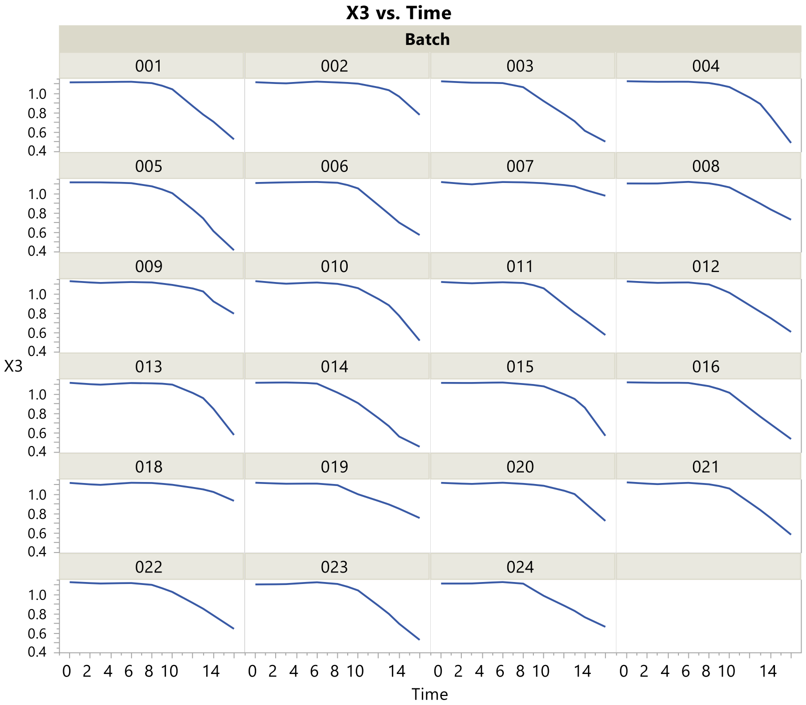
However, this isn’t always the case – and we still have 52 more variables to explore!

FUNCTIONAL DATA ANALYSIS

- Data from wearable devices will often be “telemetric” type - many repeated measures of several metrics through time.
  - This is true of data from many sources.
    - Machines output
    - Traditional time series data
    - Sensor data
    - Vibration signals
- A wide variety of specific tools and methods have been created to deal with this type of data:
  - Signal Processing
  - ARIMA Time Series
  - Partial Least Squares
  - Growth Curves via SEM
  - Mixed Models

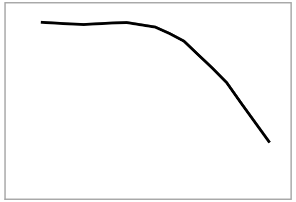
http://www.psych.mcgill.ca/misc/fda/

https://en.wikipedia.org/wiki/Functional_principal_component_analysis



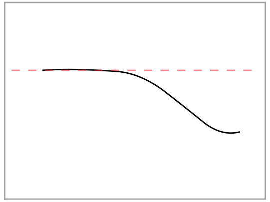
=

Mean



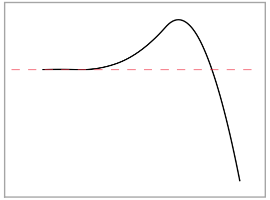
+ *w1 .*

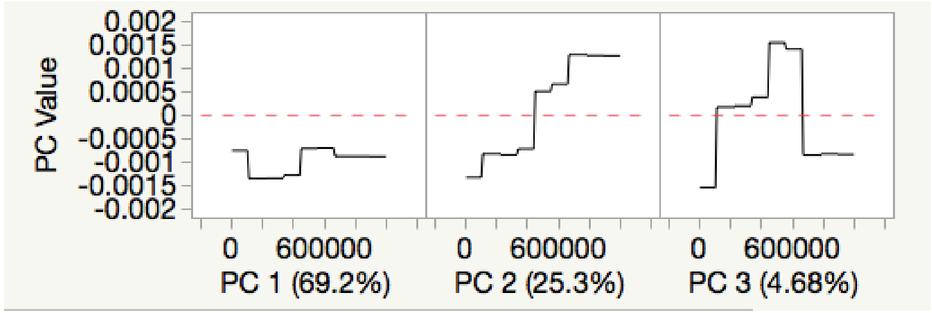
FPC1

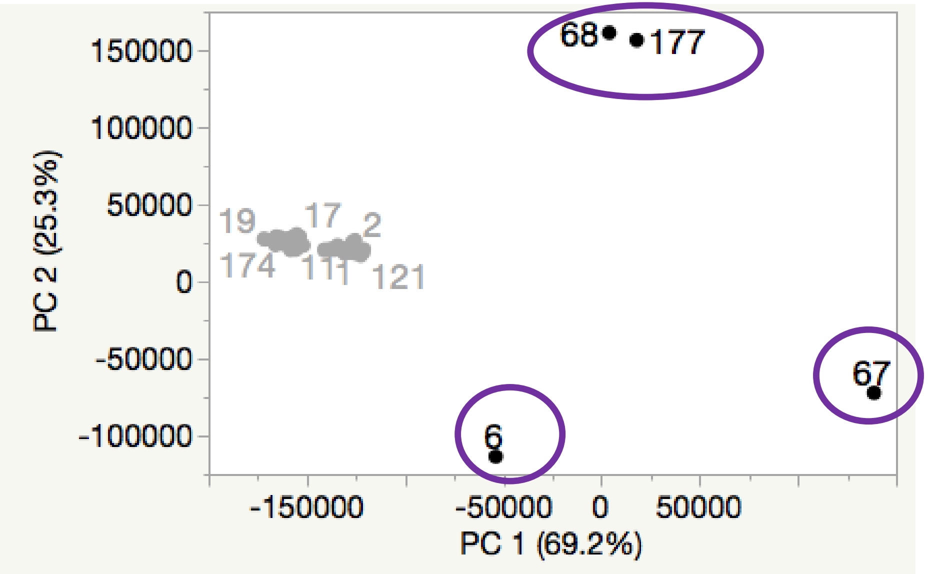


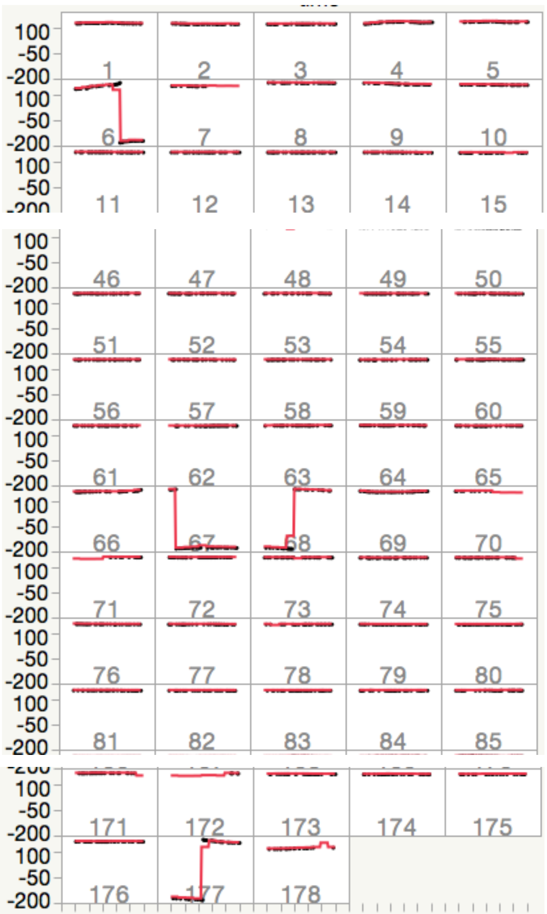
+ *w2 .*

FPC2




PC Value
0.002
0.0015
0.001
0.0005
0
-0.0005
-0.001
-0.0015
-0.002
0 600000
PC 1 (69.2%)
PC 2 (25.3%)
PC 3 (4.68%)

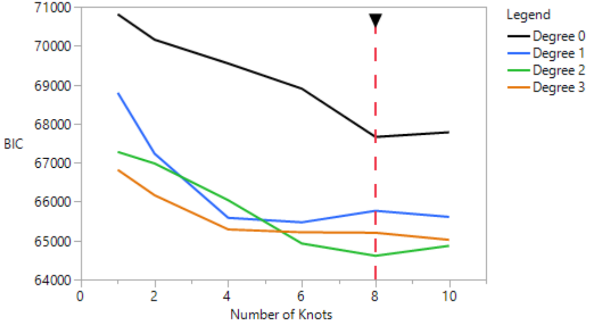

PC 2 (25.3%)
150000
100000
50000
0
-50000
-100000
-150000
-50000
0
50000
PC 1 (69.2%)
68
177
19
17
2
174
111
121
67
6

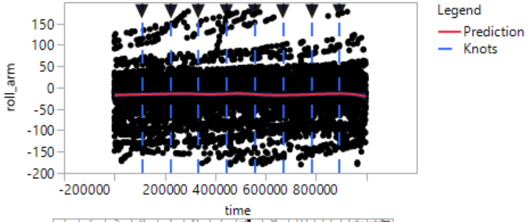

100
-50
-200
1 2 3 4 5
6 7 8 9 10
11 12 13 14 15
46 47 48 49 50
51 52 53 54 55
56 57 58 59 60
61 62 63 64 65
66 67 68 69 70
71 72 73 74 75
76 77 78 79 80
81 82 83 84 85
171 172 173 174 175
176 177 178

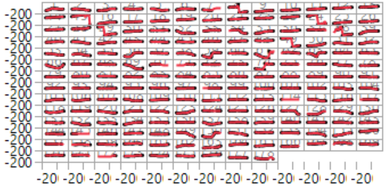
FUNCTIONAL DATA ANALYSIS

B-Spline on Initial data

Model Selection



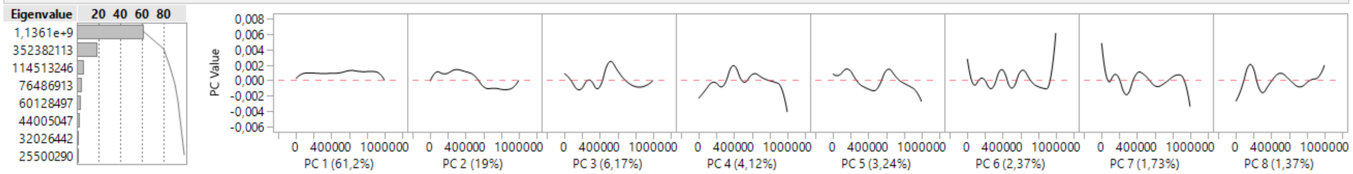




Function Summaries

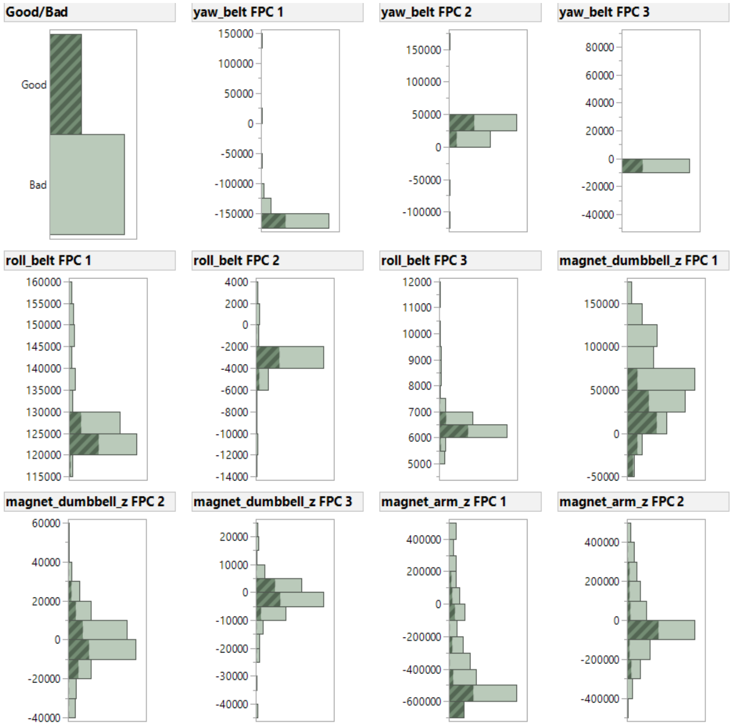
Rep	FPC 1	FPC 2	FPC 3
1	-45887,67	-623,9601	20189,664
2	-52098,08	21964,222	8894,5896
3	-65697,18	-20440,98	5663,2596
4	-38462,94	4232,6954	364,72936
5	-61173,04	24871,184	-4734,054
6	-39795,43	2761,4456	11099,398
7	-55468,97	-21449,6	-4325,925

Functional PCA

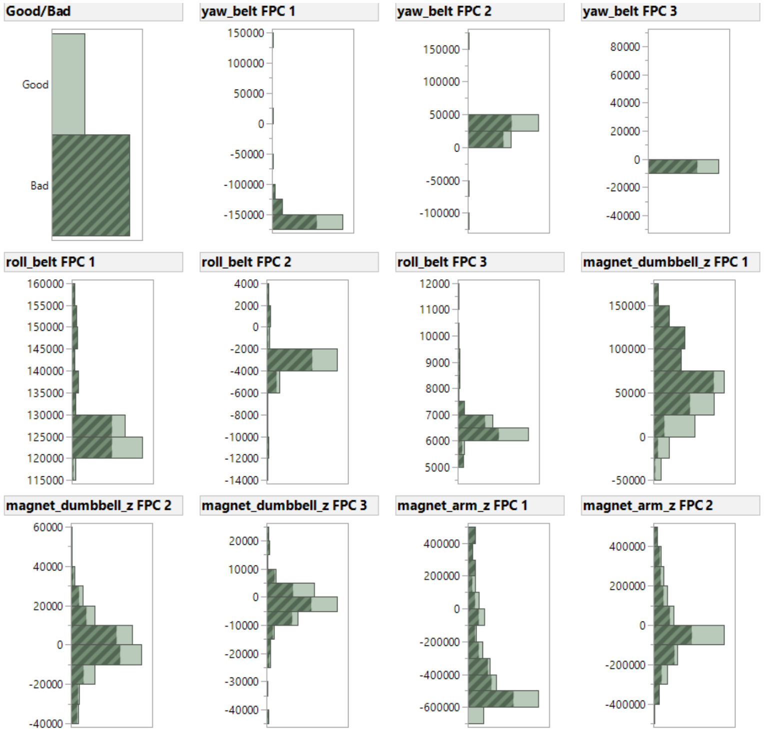


Functional Data Analysis modeling for determining functional principal components to be used in prediction.

EXPLORATORY DATA VISUALIZATION

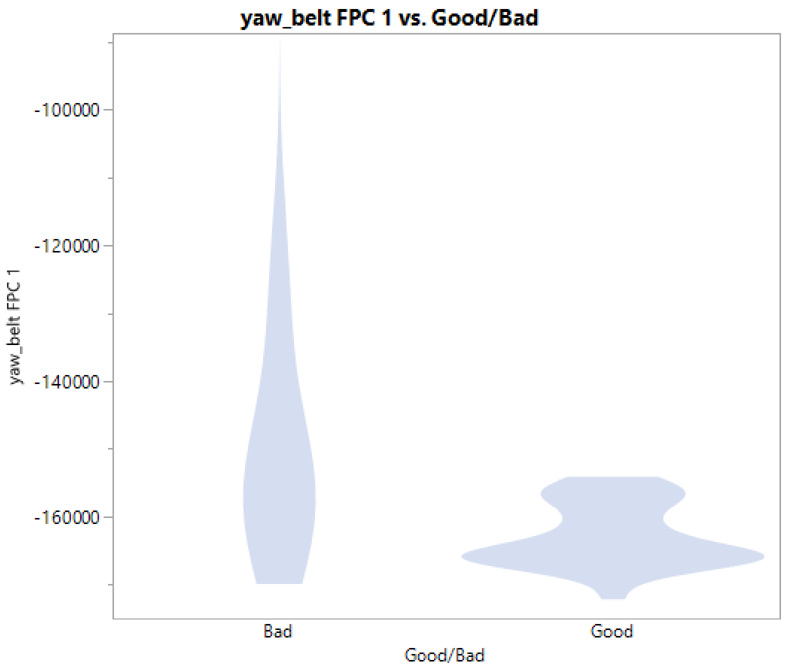


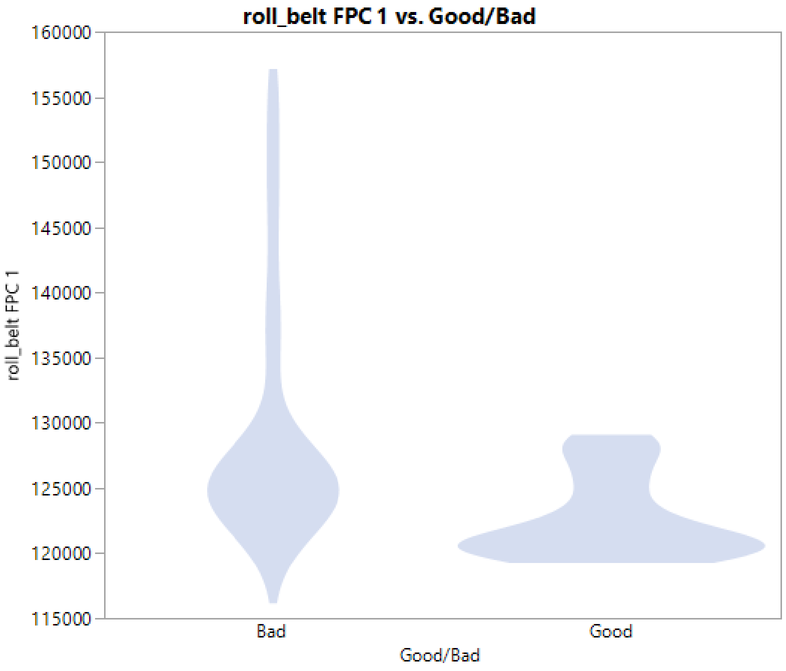
Histogram showing distribution of Good bicep curls



Histogram showing distribution of Bad bicep curls

EXPLORATORY DATA VISUALIZATION





Violin plot of Functional Principal Components vs Bad/Good

PREDICTIVE MODELING

Overall Statistics

Measure	Training	Validation	Test	Definition
Entropy RSquare	0,7890	0,5008	0,5965	1-Loglike(model)/Loglike(0)
Generalized RSquare	0,8809	0,6215	0,7138	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0,1354	0,2459	0,2077	∑ -Log(ρ[j])/n
RMSE	0,1511	0,2468	0,2261	√ ∑(y[j]-ρ[j])^2/n
Mean Abs Dev	0,1216	0,2041	0,1718	∑ |y[j]-ρ[j]|/n
Misclassification Rate	0,0000	0,0556	0,0526	∑ (ρ[j]≠ρMax)/n
N	123	36	19	n

Confusion Matrix

Training

Actual Good/Bad	Predicted Count Bad	Predicted Count Good
Bad	81	0
Good	0	42

Validation

Actual Good/Bad	Predicted Count Bad	Predicted Count Good
Bad	29	0
Good	2	5

Test

Actual Good/Bad	Predicted Count Bad	Predicted Count Good
Bad	15	0
Good	1	3

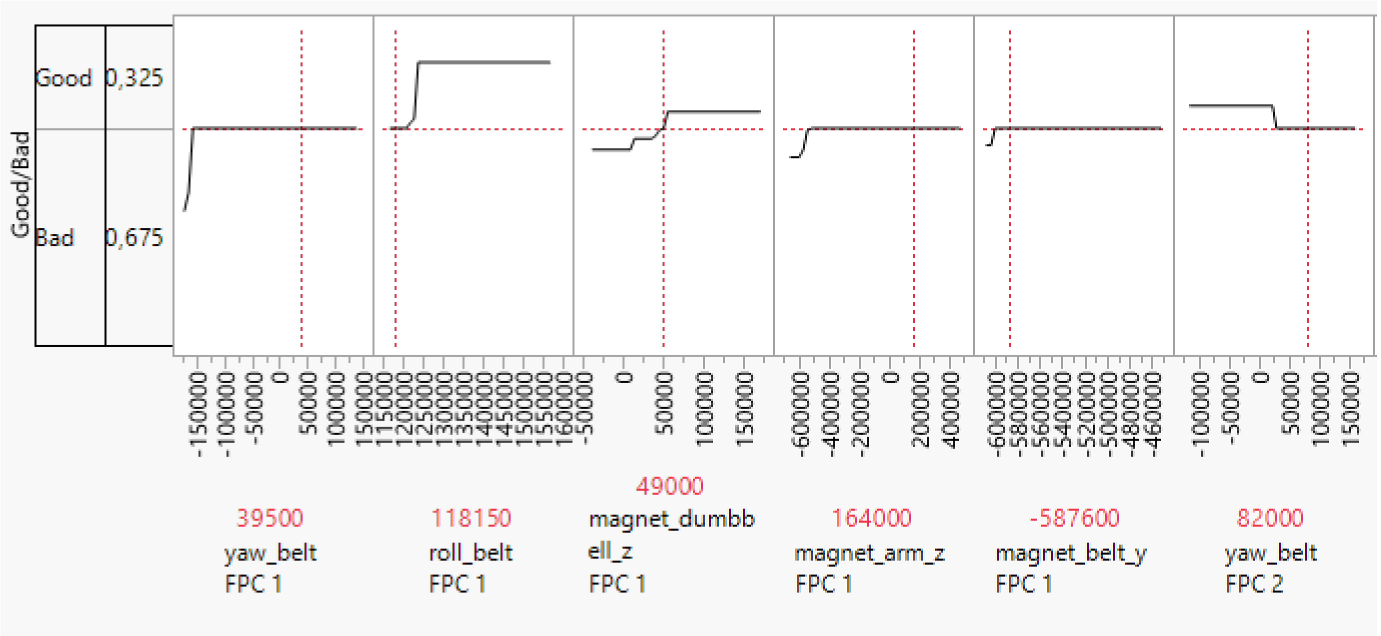
Predictive modeling results using Random Forest

Column Contributions

Term	Number of Splits	G^2		Portion
yaw_belt FPC 1	18	12,4292884		0,1411
roll_belt FPC 1	12	10,9417496		0,1242
magnet_dumbbell_z FPC 1	14	7,56536005		0,0859
magnet_arm_z FPC 1	15	6,96169266		0,0790
magnet_belt_y FPC 1	7	3,43609159		0,0390
yaw_belt FPC 2	5	2,89458554		0,0329
magnet_arm_x FPC 1	8	2,50359549		0,0284
magnet_forearm_x FPC 3	7	2,42488651		0,0275
accel_belt_y FPC 1	5	2,27927422		0,0259
magnet_forearm_z FPC 1	6	2,01215381		0,0228
accel_forearm_z FPC 1	5	1,98211032		0,0225

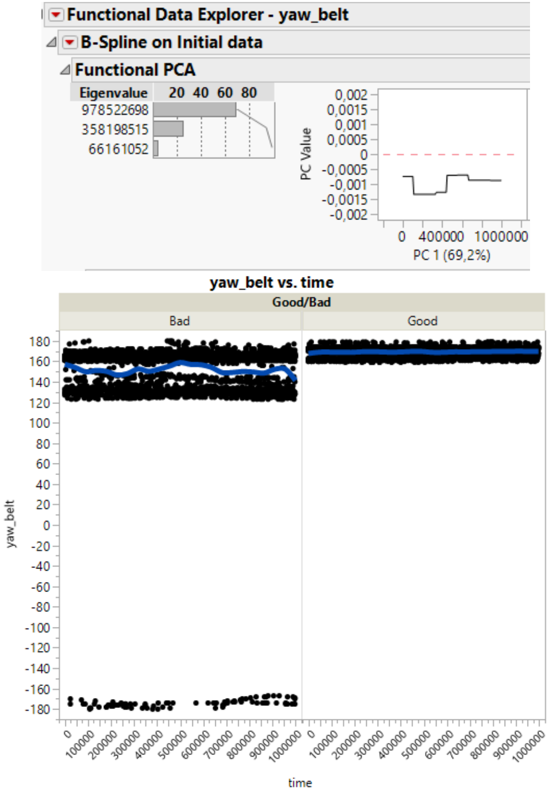
Top summary column contributors to RF model

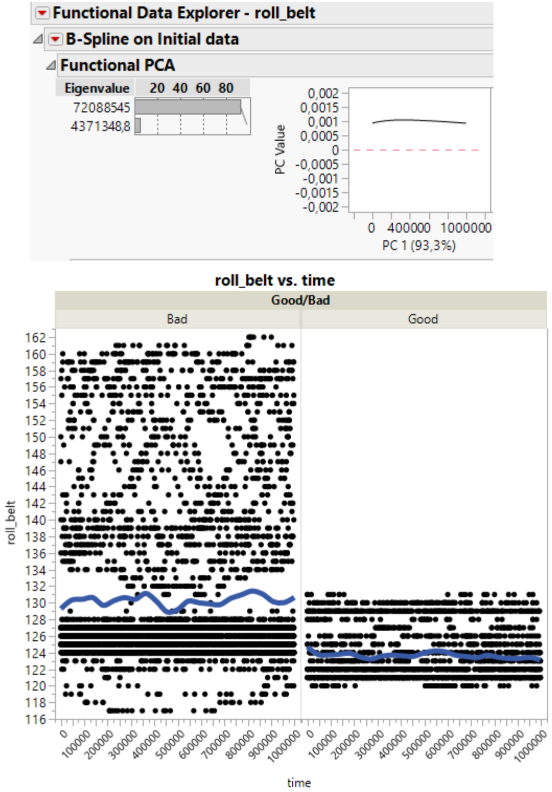
ENGAGEMENT OF STAKEHOLDERS

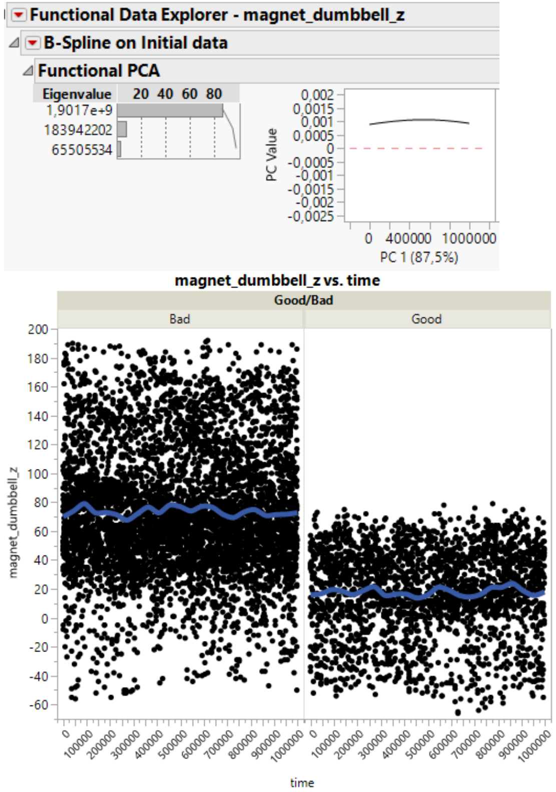


Prediction Profiler for visualizing and understanding complex models

ENGAGEMENT OF STAKEHOLDERS







Referring back to Functional PCA and time-series plots for interpretation of model.

SCALE-UP AND AUTOMATION

Formula Depot

Manage Models and Generate Scoring Code

Formula Scripts

Fit Generalized/Ridge - Good/Bad

Publish Prediction Formula

Table	Workout Data_continuous variables- (user_name = adelmo) By (num_window, Good-Bad)
Creator	Fit Generalized
Target	Good/Bad
Created	Most Likely Good/Bad
Factors	accel_arm_x Maximum[num_window], accel_arm_z Standard Deviation[num_window], magnet_forearm_z FPC 6[num_window], magnet_forearm_z Maximum[num_window], roll_belt FPC 1[num_window], roll_belt Integrated Difference[num_window], total_accel_arm Mean[num_window]
Others	Probability(Good/Bad=Bad), Probability(Good/Bad=Good)

Python Code Output - JMP

File Edit Tables DOE Analyze Graph Six Sigma Tools Tools Add-Ins View Window Help

Workout Data_continuou

```
from __future__ import division
from math import *
import jmp_score as jmp


""" ==================================================================
 Copyright(C) 2018 SAS Institute Inc.All rights reserved.

 Notice:
 The following permissions are granted provided that the
 above copyright and this notice appear in the score code and
 any related documentation. Permission to copy, modify
 and distribute the score code generated using
 JMP(R) software is limited to customers of SAS Institute Inc. ("SAS")
 and successive third parties, all without any warranty, express or
 implied, or any other obligation by SAS. SAS and all other SAS
 Institute Inc. product and service names are registered
 trademarks or trademarks of SAS Institute Inc. in the USA
 and other countries. Except as contained in this notice,
 the name of the SAS Institute Inc. and JMP shall not be used in
 the advertising or promotion of products or services without
 prior written authorization from SAS Institute Inc.
 ================================================================== """

""" Python code generated by JMP v14.0.0 """

def getModelMetadata():
    return {"creator": u"Fit Generalized", "modelName": u"Ridge", "predicted": u"Good/Bad", "tal


def getInputMetadata():
    return {
        u"accel_arm_x Maximum[num_window]": "float",
        u"accel_arm_z Standard Deviation[num_window]": "float",
        u"magnet_forearm_z FPC 6[num_window]": "float",
        u"magnet_forearm_z Maximum[num_window]": "float",
        u"roll_belt FPC 1[num_window]": "float",
        u"roll_belt Integrated Difference[num_window]": "float",
        u"total_accel_arm Mean[num_window]": "float"
    }


def getOutputMetadata():
    return {
        u"Probability( Good/Bad=Good )": "float",
```

SAS Window - JMP

File Edit Tables DOE Analyze Graph Six Sigma Tools Tools Add-Ins View Window Help

Workout Data_continuou

```
/*
 * SAS code generated by JMP v14.0.0
 * From the following model(s):
 * Fit Generalized/Ridge - Good/Bad
 */
DECLARE int idx_max;
DECLARE double val_max;
DECLARE double acc_max;
idx_max = 0;
acc_max = .;

'Probability( Good/Bad=Good )'n = (1/(1+exp(-((-793,356658153349 + -0,0234286110508104*'accel_ar
  -0,1050960584474 5*'accel_arm_z Standard Deviation[num_window]'n + -0,0001177720378326 22*'magne
  -0,052692280029945 5*'magnet_forearm_z Maximum[num_window]'n + 0,0189520946210257*'roll_belt FF
  'roll_belt Integrated Difference[num_window]'n + 4,31648257734446*'total_accel_arm Mean[num_wi
  'accel_arm_x Maximum[num_window]'n**2 + 0,000000652572785357 8*'accel_arm_x Maximum[num_window]
  'accel_arm_z Standard Deviation[num_window]'n**2 + -0,000000693576934989 6*'accel_arm_z Standar
  3 + 9,99643345901826e-11*'magnet_forearm_z FPC 6[num_window]'n**2 + 2,26564050057371e-14*'magr
  3 + 0,00008655163850501 85*'magnet_forearm_z Maximum[num_window]'n**2 + -0,00000005183943402 46*
  3 + -0,000000149451048337 1*'roll_belt FPC 1[num_window]'n**2 + 3,78018291463241e-13*'roll_belt
  'roll_belt Integrated Difference[num_window]'n**2 + 0,00038060274951031 5*'roll_belt Integrated
  3 + -0,17007422595295 4*'total_accel_arm Mean[num_window]'n**2 + 0,0017882603322208 6*'total_acc
  3))))));
'Probability( Good/Bad=Bad )'n = 1 + -1*(1/(1+exp(-((-793,356658153349 + -0,0234286110508104*'ac
  -0,1050960584474 5*'accel_arm_z Standard Deviation[num_window]'n + -0,0001177720378326 22*'magne
  -0,052692280029945 5*'magnet_forearm_z Maximum[num_window]'n + 0,0189520946210257*'roll_belt FF
  'roll_belt Integrated Difference[num_window]'n + 4,31648257734446*'total_accel_arm Mean[num_wi
  'accel_arm_x Maximum[num_window]'n**2 + 0,000000652572785357 8*'accel_arm_x Maximum[num_window]
  'accel_arm_z Standard Deviation[num_window]'n**2 + -0,000000693576934989 6*'accel_arm_z Standar
  3 + 9,99643345901826e-11*'magnet_forearm_z FPC 6[num_window]'n**2 + 2,26564050057371e-14*'magr
  3 + 0,00008655163850501 85*'magnet_forearm_z Maximum[num_window]'n**2 + -0,00000005183943402 46*
  3 + -0,000000149451048337 1*'roll_belt FPC 1[num_window]'n**2 + 3,78018291463241e-13*'roll_belt
  'roll_belt Integrated Difference[num_window]'n**2 + 0,00038060274951031 5*'roll_belt Integrated
  3 + -0,17007422595295 4*'total_accel_arm Mean[num_window]'n**2 + 0,0017882603322208 6*'total_acc
  3))))));
val_max = 'Probability( Good/Bad=Good )'n;
IF (not missing(val_max) AND (missing(acc_max) OR val_max > acc_max)) THEN DO;
  idx_max = 1;
  acc_max = val_max;
END;
val_max = 'Probability( Good/Bad=Bad )'n;
IF (not missing(val_max) AND (missing(acc_max) OR val_max > acc_max)) THEN DO;
  idx_max = 2;
  acc_max = val_max;
END;
```

Generate code from prediction model formula.

SUMMARY

ANALYSIS, VISUALIZATION AND INTERPRETATION OF BIOMETRIC DATA FROM MOBILE SMART DEVICES

- We showed you how you can:

1. Access biometric data from smart devices.
2. Compile and organize said data.
3. Clean data.
4. Visualize and explore.
5. Create predictive models.
6. Stakeholder engagement.
7. Generate code from predictive models for scale-up activities.

Special attention may be needed to extract meaningful features from telemetric-type data.