IB3K50

Artificial Intelligence for Business

December 14, 2020

Agenda

Task 1: Assessing the benefits of AI in the Energy Industry – A look at Wind Power

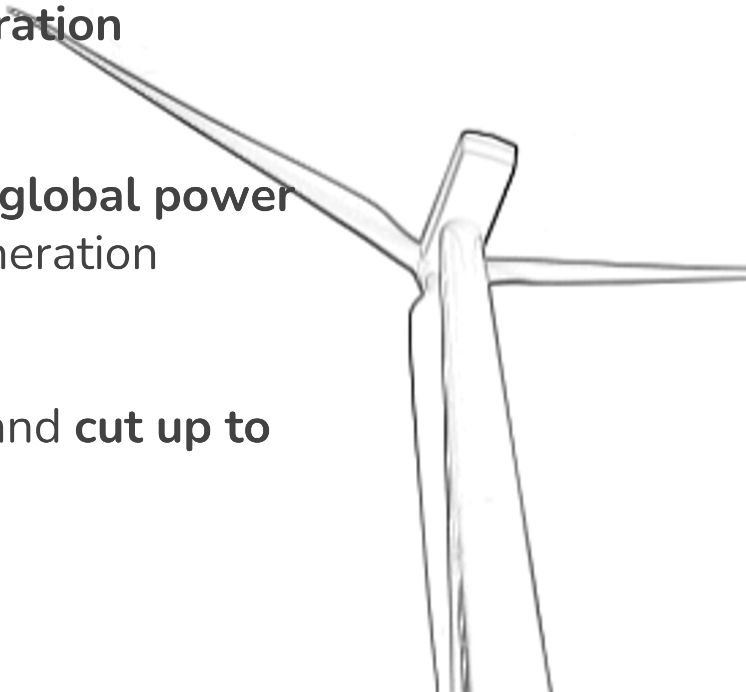
Task 2: Analysing 1 year of sports activity data and building predictive models

1. Introduction
Innovation to support the future of wind power

Increasing demand for renewable energies which are expected to become the **largest source of electricity generation worldwide by 2025.**

Wind power could cover more than **one-third of global power needs** (35%), becoming the world's foremost generation source.

Artificial intelligence input to increase efficiency and **cut up to 30% of maintenance costs.**



(IEA, 2020; IRENA 2019; Wang, KS., Sharma, V.S. & Zhang, ZY. 2014)

Task 1: Assessing the benefits of AI in the Energy Industry – A look at Wind Power

2. Value Chain Analysis
AI in the Energy Value Chain

Support Activities

Firm Infrastructure Digital Twin

Human Resources Applicant Screening System

Technology Development Digital Twin

Procurement

Inbound Logistics

- Forecasting supply
- DERs

Operations

- Smart Grid
- Digital Twin

Outbound Logistics

- Forecasting Demand
- Smart Metering
- Smart Energy Storage

Marketing and Sales

- Client Segmentation and Pricing

Service

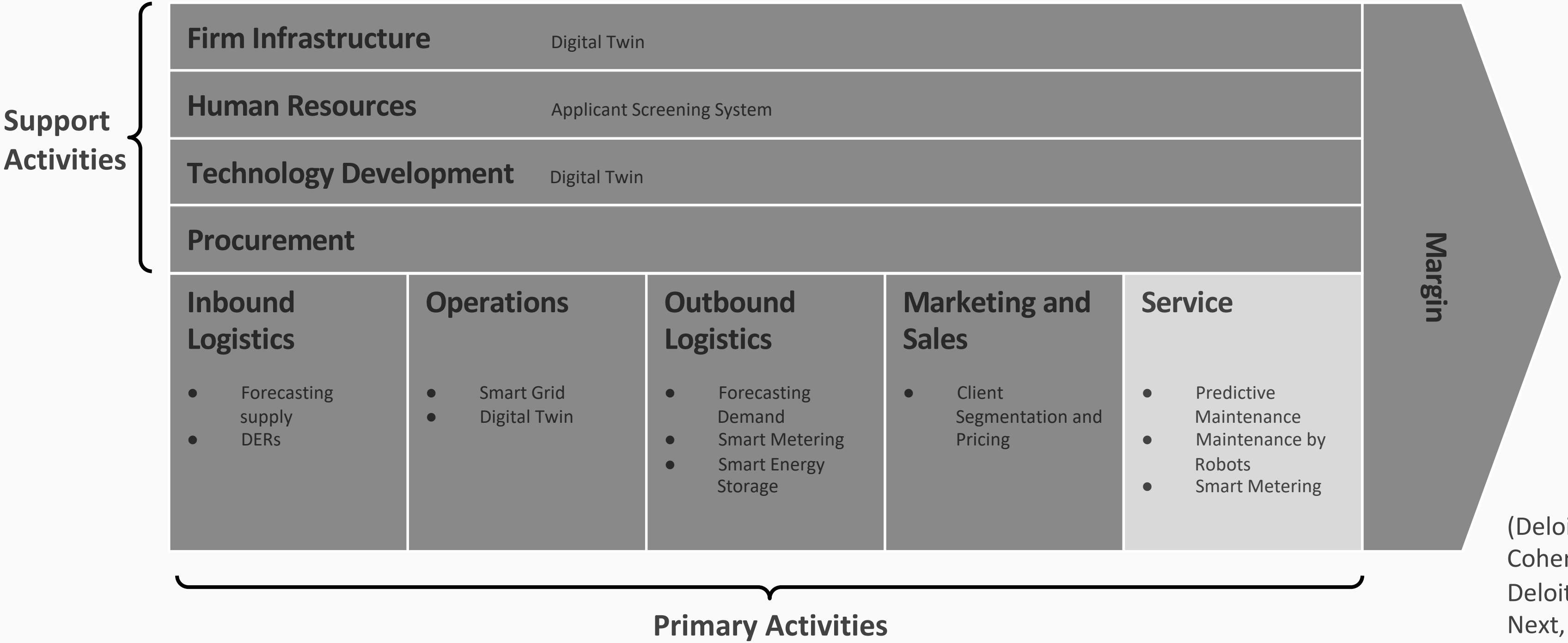
- Predictive Maintenance
- Maintenance by Robots
- Smart Metering

Margin

Primary Activities

(Deloitte, 2015; Cohen, 2020; Deloitte, 2019; Next, n.d.)

2. Value Chain Analysis
AI in the Energy Value Chain



(Deloitte, 2015; Cohen, 2020; Deloitte, 2019; Next, n.d.)

3. AI in Predictive Maintenance

From gathering data to improving business performance

DATA SOURCE	PROCESSING	BENEFITS
Wind turbine motors		Operational efficiency
SCADA data	Machine Learning	Optimised maintenance and asset management
Meteorological masts	Artificial Intelligence	Cost reduction
Power plant substation equipment		Overall business performance

Schlechtingen and Ferreira Santos, 2018; Stetco et al., 2019

3. Neural Network
Model

M. Schlechtingen, I. Ferreira Santos / Mechanical Systems and Signal Processing 25 (2011) 1849–1875

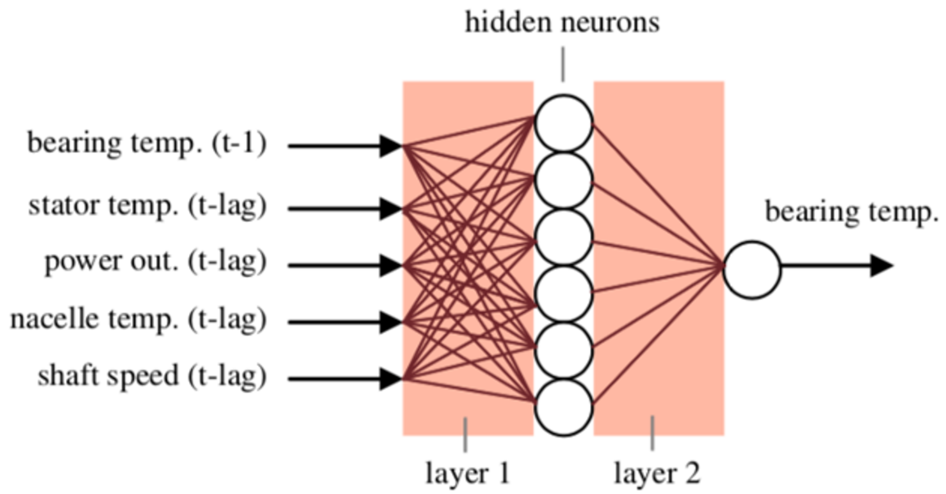


Fig. 18. Network architecture—feed forward network [6].

Schlechtingen and Ferreira Santos, 2018

3. Neural Network
Critical Evaluation

Computationally Expensive

Neural network requires more data than traditional ML algorithms

Network Architecture

Identifying the suitable network setup:

- input signals
- number of training patterns
- network structure (number of layers, neurons; connectivity)
- weight initialization
- backpropagation

Complexity of data pre-processing:

- validity check
- data scaling
- missing data processing
- lag removal

“Black Box” Nature

Why or how the model came up with a certain prediction of potential failure

Limits interpretability

Schlechtingen and Ferreira Santos, 2018; Stetco et al., 2019

3. Benefits
Vestas Example

Reactive Maintenance	Predictive Maintenance
€7,000 AND 3 DAY MINIMUM STOP	**€100 AND NO STOP**
€1,000 AND 3 DAY MINIMUM STOP	**€15 AND NO STOP**

Chen, 2017

3. Benefits
Measurable Benefits

OPERATING COST

- Reduces asset failure : cheaper to repair before machine breakdown happens
- Extends asset life and reliability
- Less workforce allocated to maintenance
- Diminishes stock-related costs

TIME SAVING

- Instant trouble-detection & first visit repair
- Reduces asset downtime: spare parts are available at the "right-place right-time"
- Cuts spare parts ordering time lag

INVENTORY OPTIMIZATION

- Defines best inventory composition regarding seasonality/specific factors
- Cuts the parts ordering process
- Better resource allocation: optimized utilisation of capacity and workforce

Mobley, 2001; Dezide, 2020; Kent, 2018

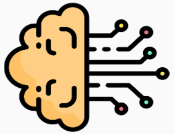
3. Benefits
Intangible Benefits

Training

Employees’ experience is centralized in the platform: every Vestas’ 8k field engineers can share and use another’s knowledge

Knowledge

Knowledge of Vestas’ products and wind farms’ environments



Safety

Detects dangerous situations
Prevents future hazards

Customer Relations

Meeting customers needs anticipating their expectations

Forsthoffer, 2011; Mobley, 2001; Dezide, 2020

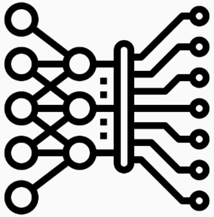
3. Critical Evaluation

Complexity

Very large amount of data to be handled

Complex separation between general and specific issues

Engineers need to be trained both for AI and for mechanics



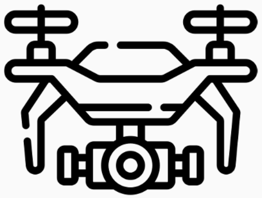
Reliability

Ecological validity of predictions is achieved through a long training of the machine on large amount of high quality data

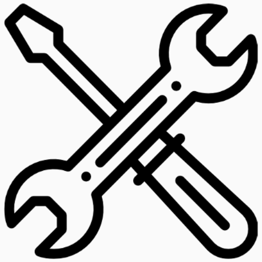
Often leads to misinterpretation of informations, thus false predictions

IBM, n.d.; Miller, 2014; Quinonero-Candela et al., 2006

3. Future Outlook



Automated Physical Inspections



Automated Physical Repairs

McKinsey Institute, 2014; Dedezade, 2019

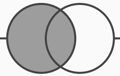
Task 2: Analysing 1 year of sports activity data and building predictive models

1. Introduction
Dataset overview

activity-data.csv

Content: Philipp’s personal activity data
Columns: 61 with dimensions (e.g. activity type) and measures (e.g. average speed)
Rows: 200 (one for each activity)

Further descriptive statistics of the data sets can be found in Appendix 3.

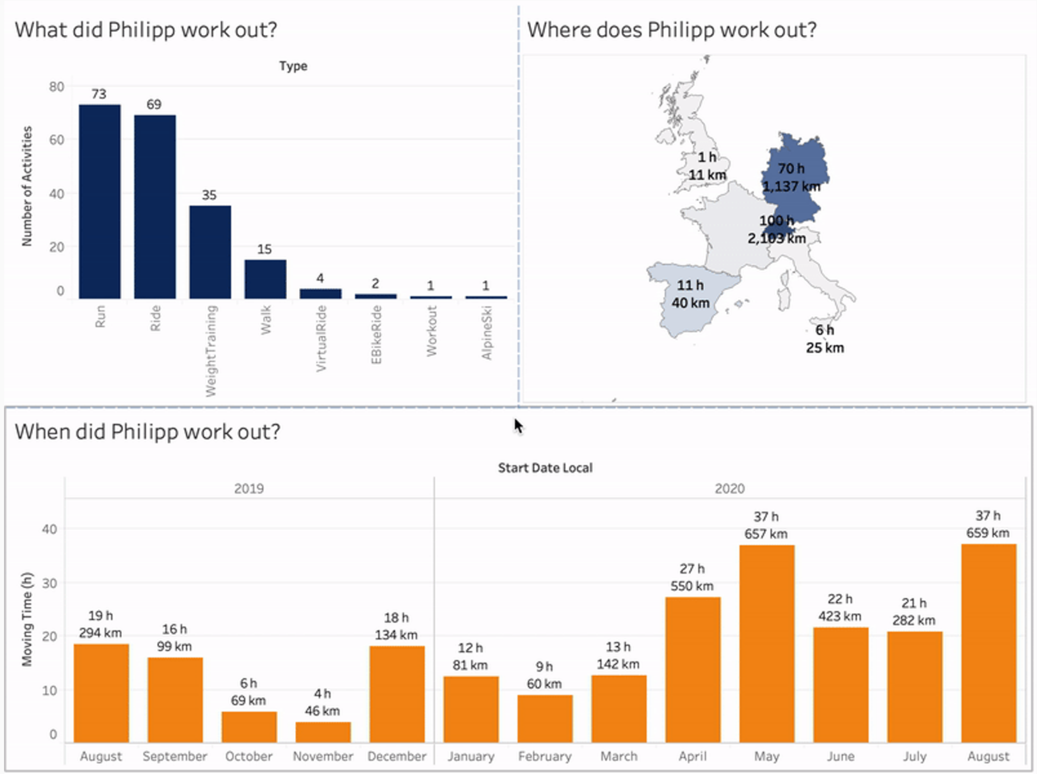


LEFT JOIN

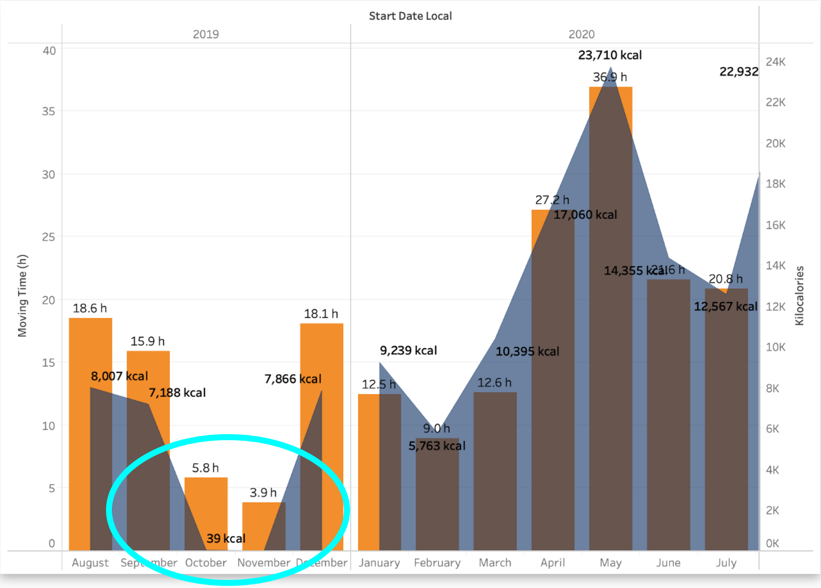
location-data.csv

Content: Data on the activity locations
Columns: 5 (e.g. activity id, city, country, etc.)
Rows: 200

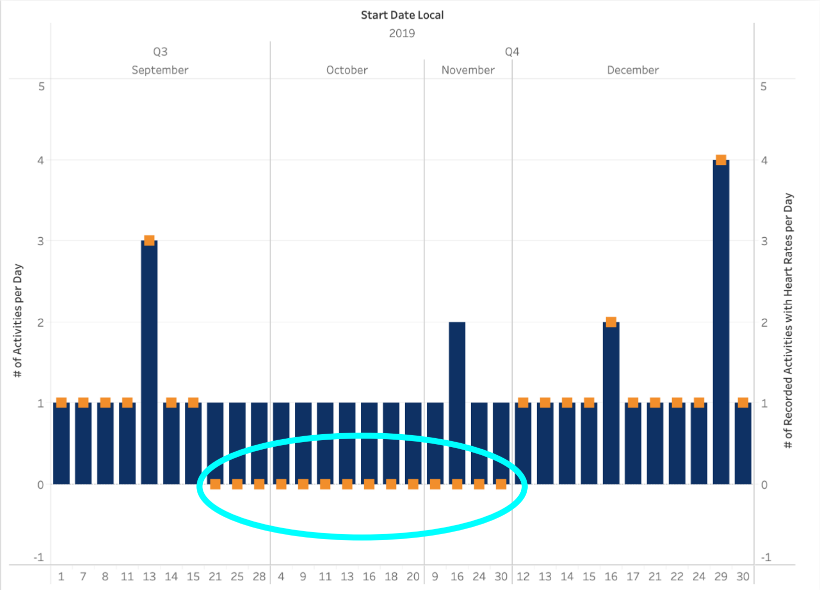
2. Interactive Data Visualisation
Dataset overview



3. Descriptive Analysis
Dataset overview

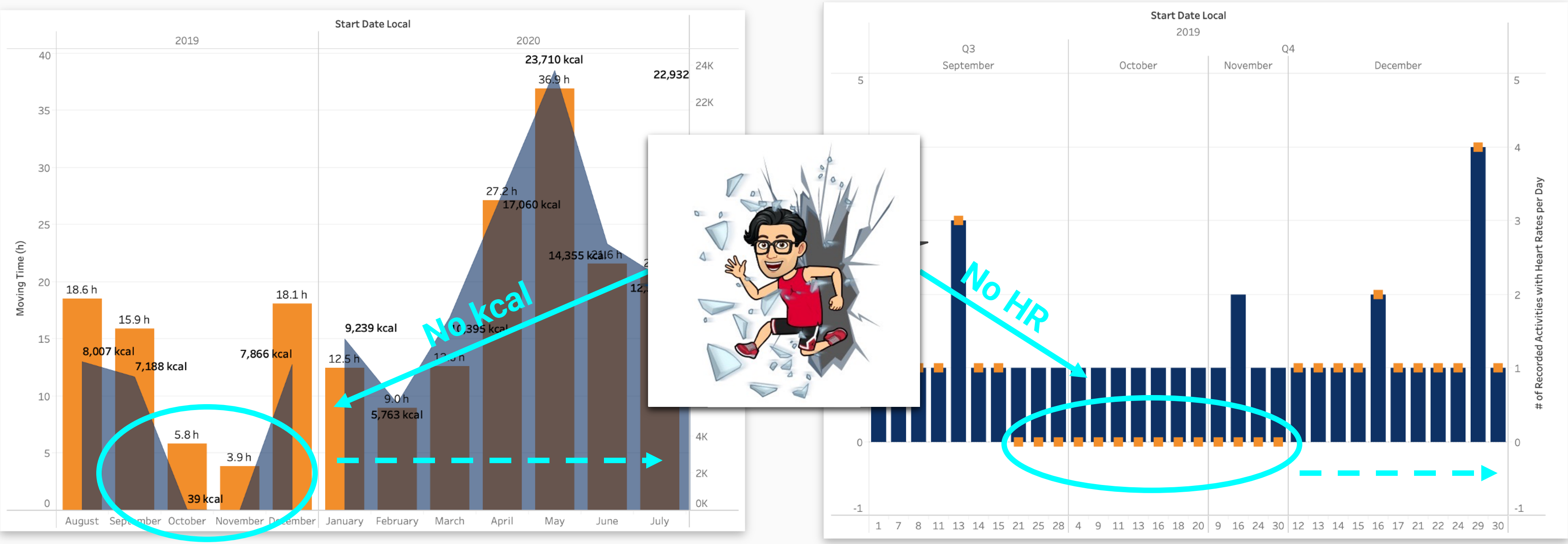


Low/no **calorie** values, despite activity



No **heart rate** recorded, despite activity

3. Descriptive Analysis
Dataset overview

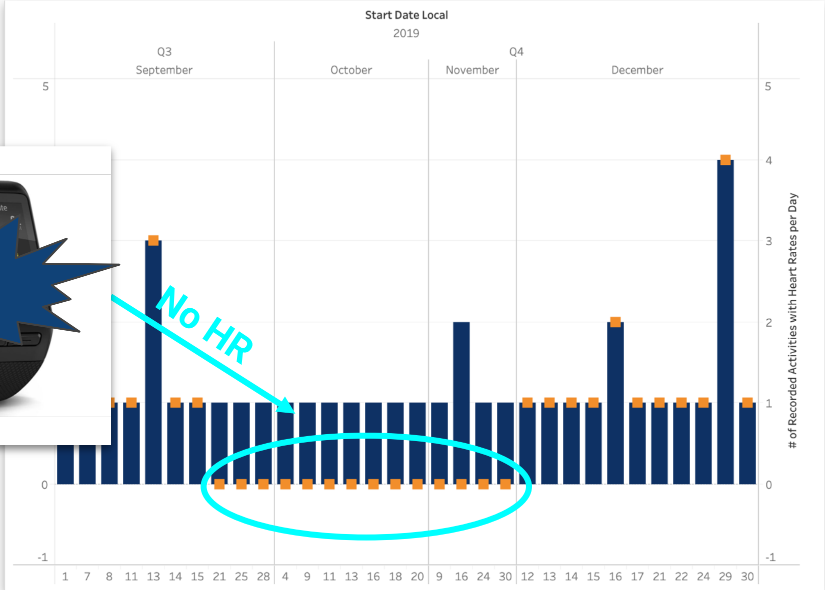


Low/no **calorie** values, despite activity

No **heart rate** recorded, despite activity

3. Descriptive Analysis
Dataset overview

Low/no **calorie** values, despite activity



No **heart rate** recorded, despite activity

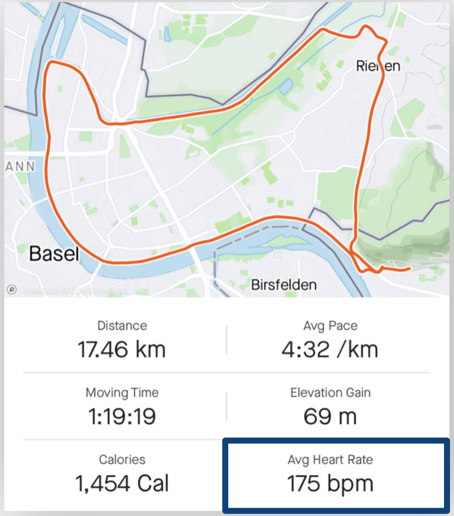
4. Linear Regression: Heart Rate “Prediction”
Rationale

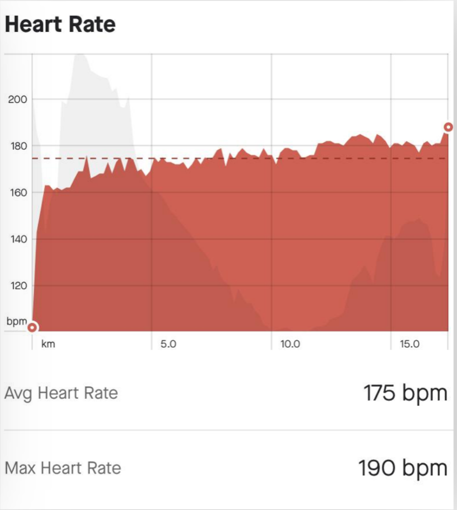
Status quo

Missing heart rate data for 14 runs in the data set

Benefit

Availability of a key measure to visualise and analyse efforts during these runs. This can help Philipp track his progress and improve his running in the long run.





4. Linear Regression: Heart Rate “Prediction” Preparation

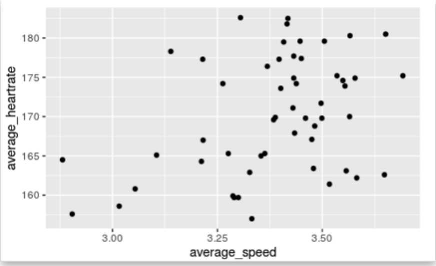
Literature review

Lambert, Mbambo and Gibson (1998) suggest a strong correlation between **running speed** and **heart rate**

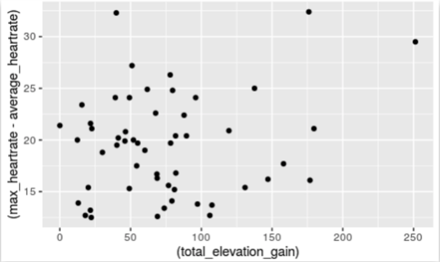
Padulo et al. (2013) suggest a correlation between **heart rate variability** and **slope**

Analysis of data set*

Very slightly positive correlation

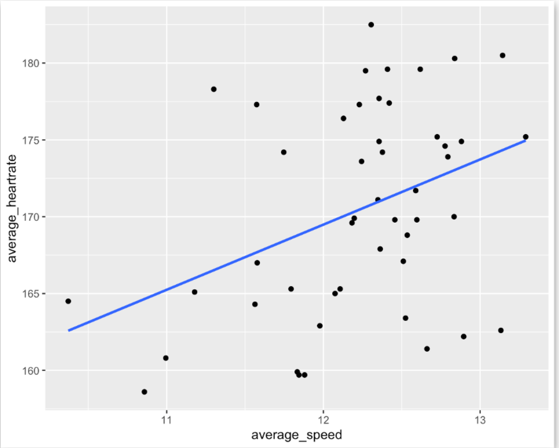


No clear correlation



*We analysed relationship between the avg. heart rate (dependent variable) and other potential independent variables too (see Appendix 1), none of them seemed to be relevant

4. Linear Regression: Heart Rate "Prediction"
Model & Utilisation

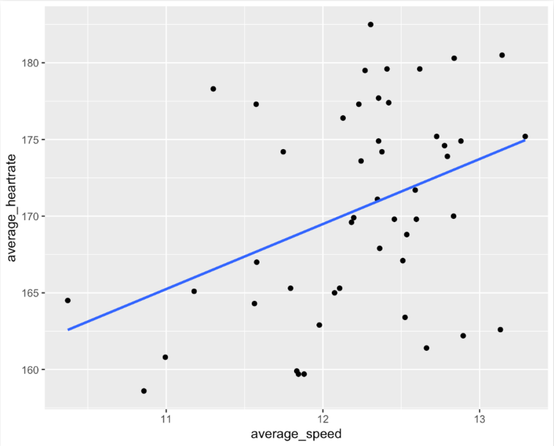



```
Residuals:
     Min       1Q   Median       3Q      Max
-11.6895  -4.4530  -0.3522   5.8744  11.7811

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    118.598     18.540   6.397 7.34e-08 ***
average_speed    4.241      1.515   2.799  0.00747 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.322 on 46 degrees of freedom
  (20 observations deleted due to missingness)
Multiple R-squared:  0.1455,    Adjusted R-squared:  0.1269
F-statistic: 7.834 on 1 and 46 DF,  p-value: 0.007467
```

has_heartrate	average_heartrate
FALSE	171.1338116
FALSE	168.5535872
FALSE	164.2786592
FALSE	171.5613044
FALSE	168.5383196
FALSE	165.4847996
FALSE	166.2939824
FALSE	166.30925
FALSE	168.0802916
FALSE	171.5613044
FALSE	173.4239516
FALSE	164.1259832
FALSE	169.5612488
FALSE	168.752066

Utilisation of the model to prepare data in excel for the visualisations

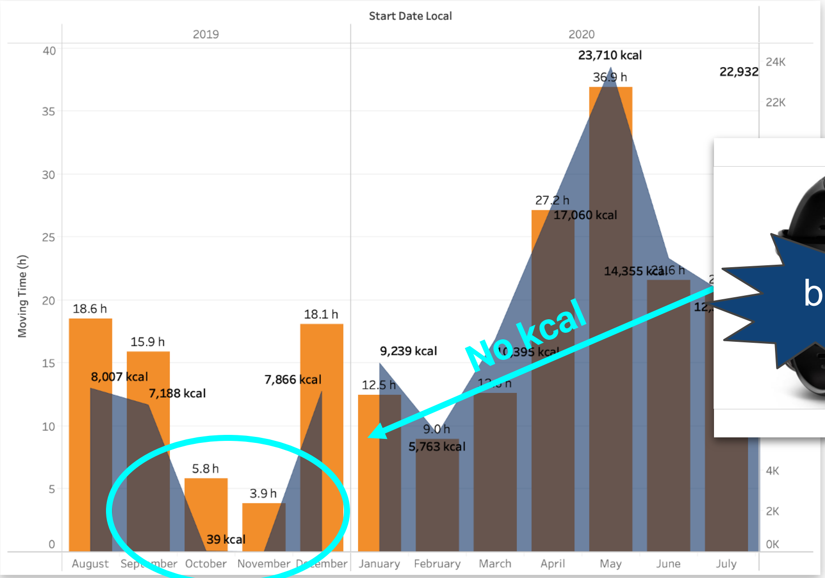
4. Linear Regression: Heart Rate "Prediction"
Critical Evaluation

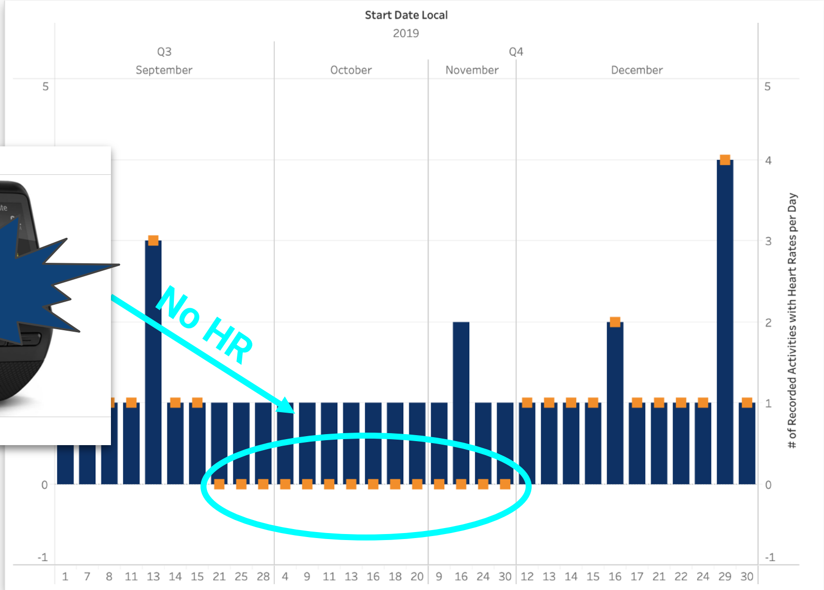



```
Residuals:
     Min       1Q   Median       3Q      Max
-11.6895  -4.4530  -0.3522   5.8744  11.7811

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    118.598     18.540   6.397 7.34e-08 ***
average_speed    4.241      1.515   2.799  0.00747 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.322 on 46 degrees of freedom
  (20 observations deleted due to missingness)
Multiple R-squared:  0.1455,	Adjusted R-squared:  0.1269
F-statistic: 7.834 on 1 and 46 DF,  p-value: 0.007467
```

- Statistically significant p-values, yet, **low R-squared** indicates that the model explains only a low percentage of the variance in average heart rate
- Trade-off between large and highly suitable training data set
  - **Accuracy**: Data set ranges over a long time period (> 10 months) with different training patterns and fitness conditions
  - **Sample Size**: Relatively larger sample size of n = 68 runs could be used for the linear regression
- Model specific to Philipp - **not generalisable**

4. Linear Regression: Heart Rate "Prediction"
Applying model output to dataset

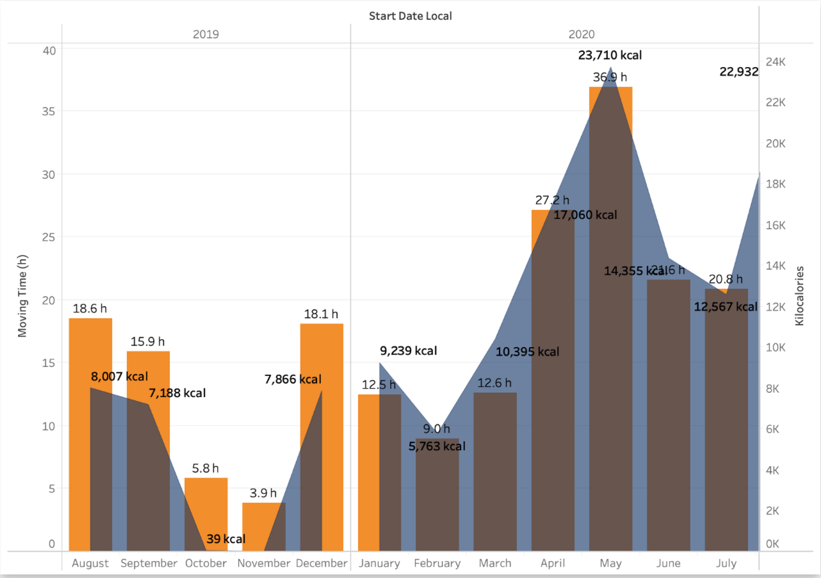


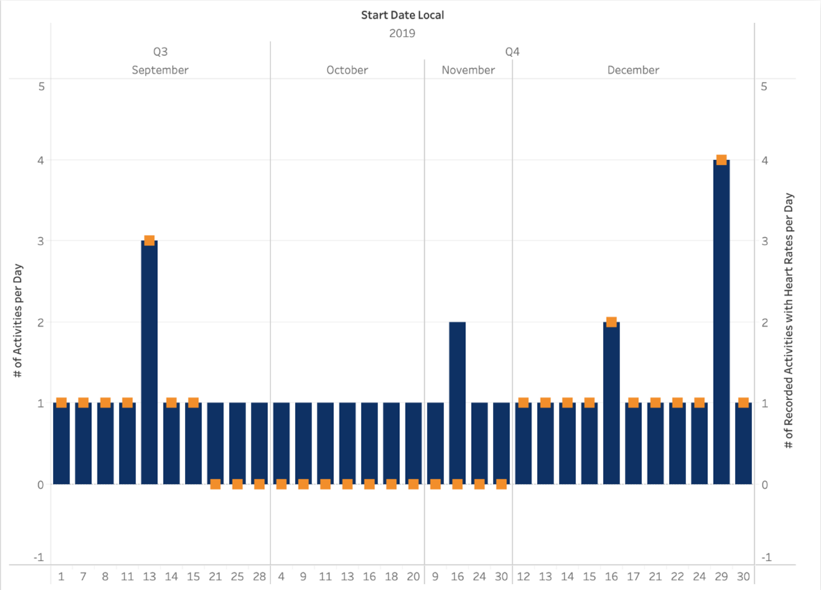


Low/no **calorie** values, despite activity

No **heart rate** recorded, despite activity

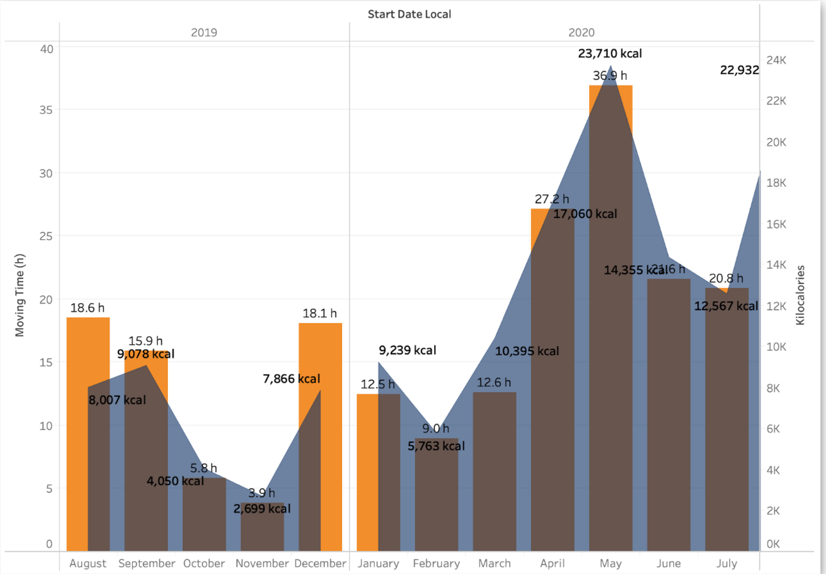
4. Linear Regression: Heart Rate “Prediction”
Applying model output to dataset

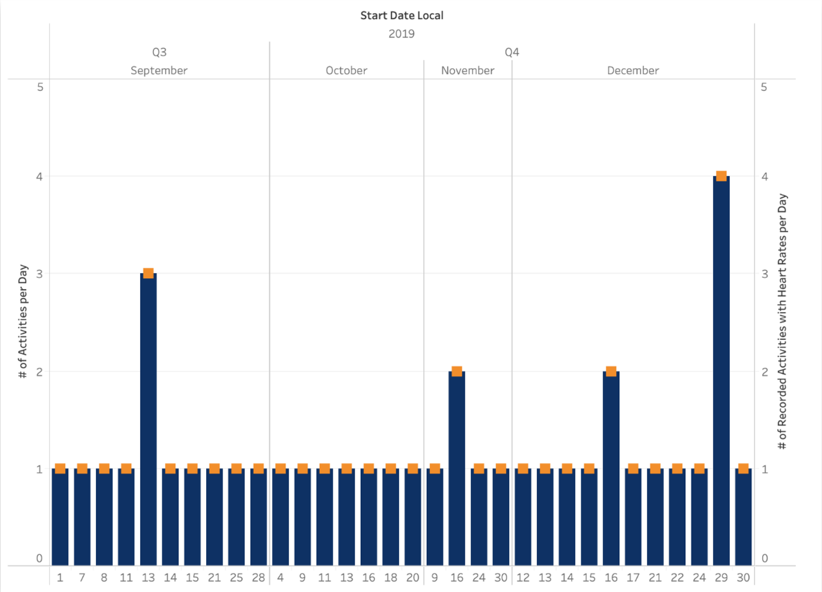




4. Linear Regression: Heart Rate “Prediction”

Applying model output to dataset





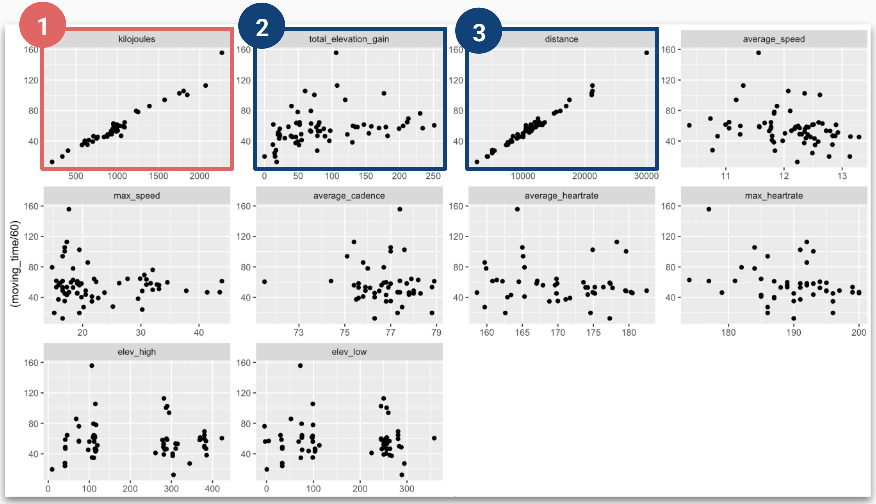
4. Linear Regression: Run Time Prediction
Preparation

Identification of independent variables*

1. Kilojoules is positively correlated with moving time, but not a value that is available before a run

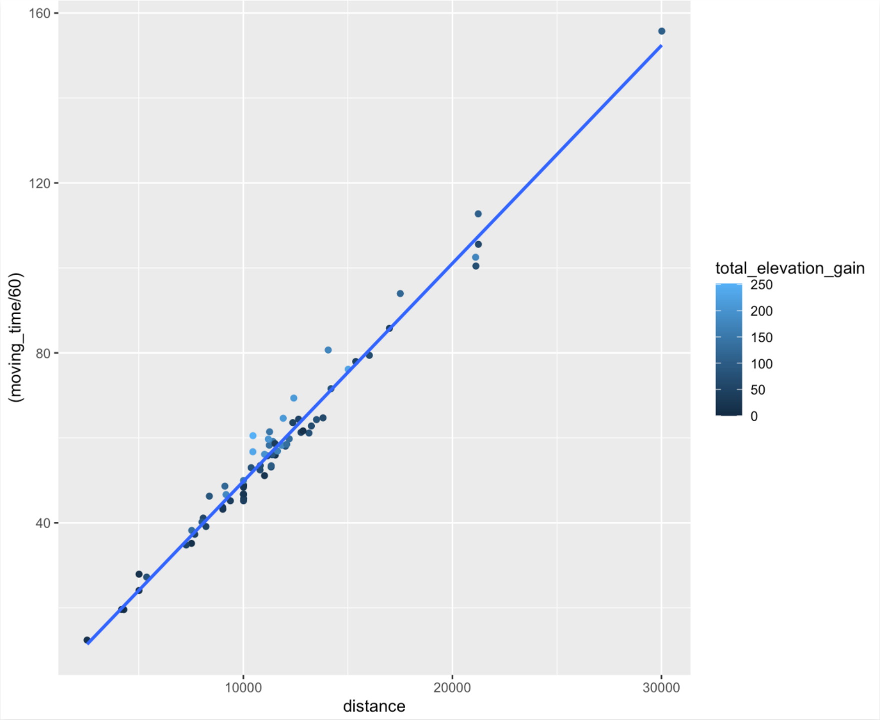
2. Total elevation gain has a very slight positive correlation with moving time and can be obtained before a run e.g. via Google Maps

3. Distance has a stronger positive correlation with moving time and can be obtained before a run e.g. via Google Maps



*Examination of correlations and potential multicollinearity is in Appendix 2

4. Linear Regression: Run Time Prediction Model




```
Residuals:
    Min      1Q  Median      3Q     Max 
-5.7484 -1.9757  0.0595  1.4379  7.8625 

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)    
(Intercept)         -2.8085563  0.9437897  -2.976  0.00401 ** 
distance             0.0050250  0.0000774  64.924  < 2e-16 ***
total_elevation_gain 0.0282762  0.0052249   5.412 8.25e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.74 on 70 degrees of freedom
Multiple R-squared:  0.9854,	Adjusted R-squared:  0.985 
F-statistic:  2363 on 2 and 70 DF,  p-value: < 2.2e-16
```

4. Linear Regression: Run Time Prediction
Critical Evaluation

- Realistically running time **won't be a linear function** (e.g. exhaustion)
- **Other factors** influence the running time, which our data set does not account for (e.g. level of effort, differences in circumstances etc.)
- Model doesn't account for **progress in training/fitness**, because it will always be trained with past training data and is only specific to Philipp
- Patterns indicated by the model do **not necessarily imply correlation** (so just a pattern), as measurements are not independent of each other
- Model indicates that if your distance and total elevation gain are 0, moving time is -3 minutes
- Data for distance and elevation gain has not been **normalised** nor **standardised**

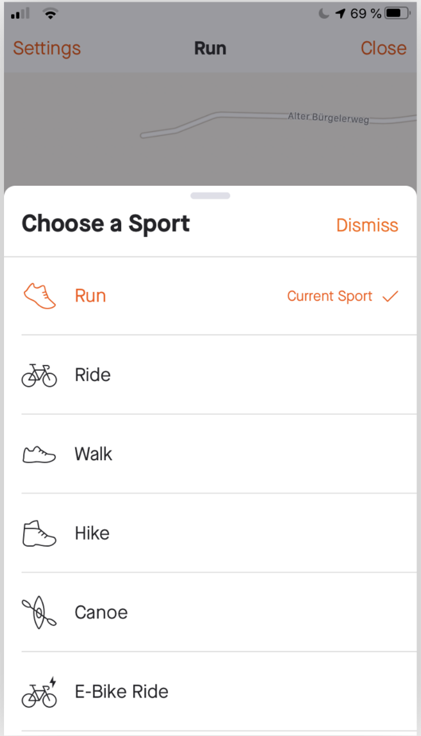
4. K-Nearest Neighbors (KNN)
Rationale

Status quo

Manual activity selection by user before activity start.

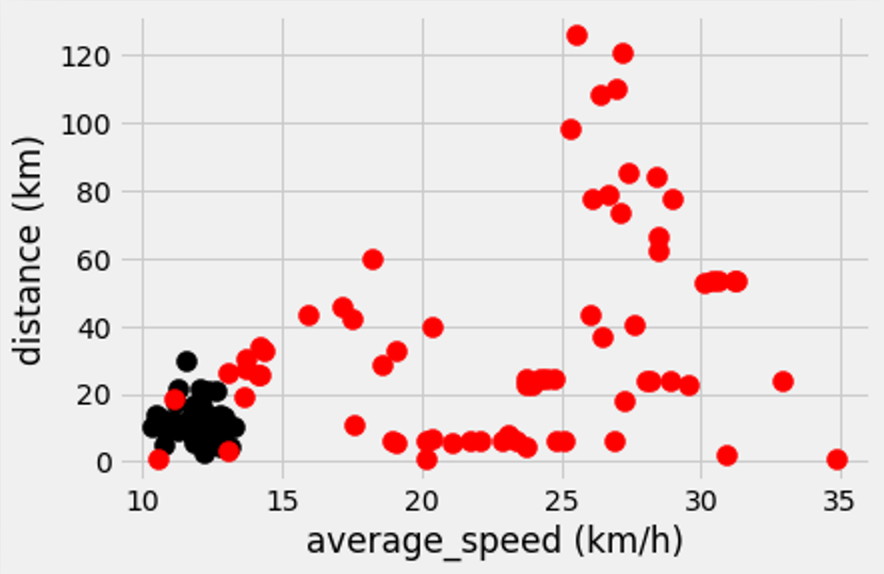
Benefit

Predict activity type based on all data gathered from the training. KNN model based on past training data.

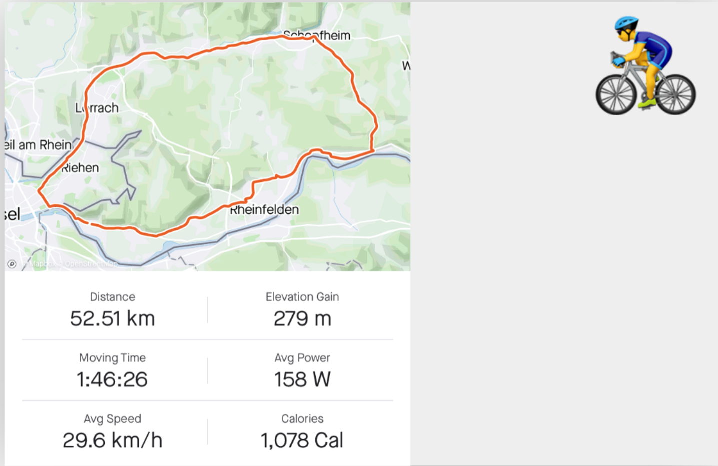


4. K-Nearest Neighbors (KNN)
Model setup

- **Training data**
  Runs: 73
  Rides: 69
- **X: Average speed (km/h)**
- **Y: Distance (km)**
- **K = 3**



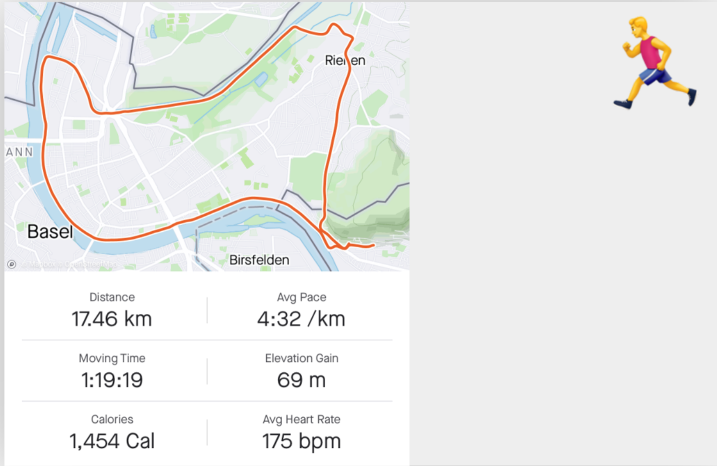
4. K-Nearest Neighbors (KNN)
Model testing



Input: [29.6, 52.51]

Average speed 29.6 km/h

Distance 52.51 km

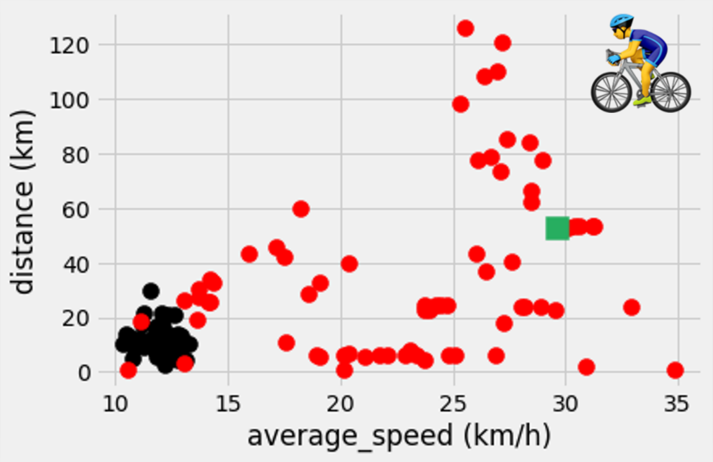


Input: [13.1, 17.46]

Average speed 13.1 km/h

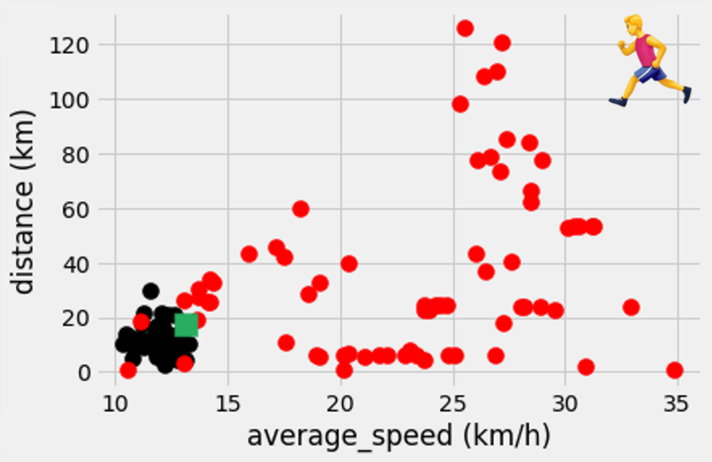
Distance 17.46 km

4. K-Nearest Neighbors (KNN)
Model testing



Input: [29.6, 52.51]

Output: Ride ✓



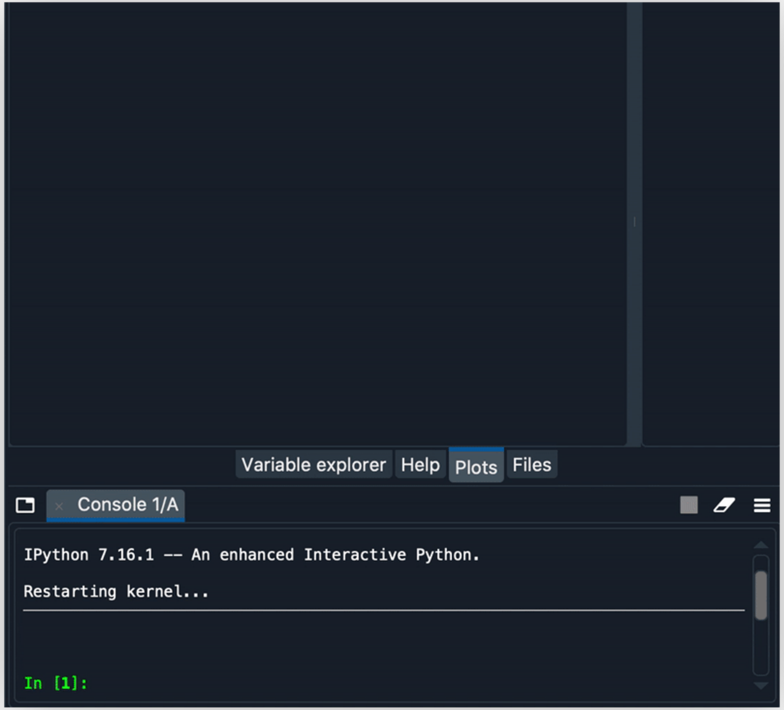
Input: [13.1, 17.46]

Output: Run ✓

4. K-Nearest Neighbors (KNN)
Model testing (extended)

k	accuracy
1	92.86%
2	92.86%
3	**95.24%**
5	95.24%
10	95.24%
40	92.86%

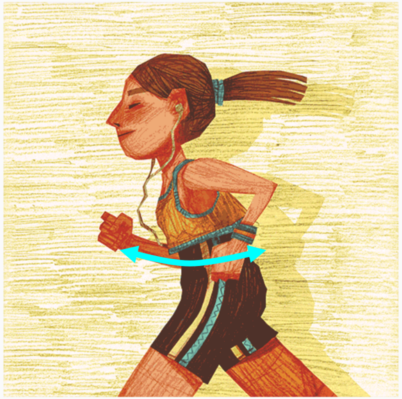
with n = 42 runs



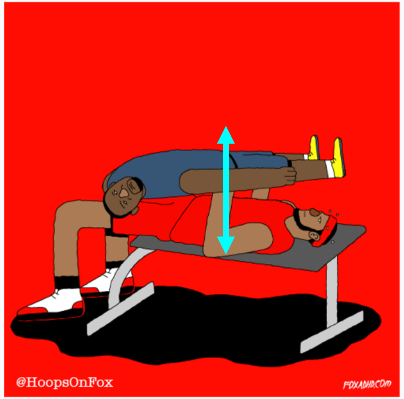
4. K-Nearest Neighbors (KNN)
Improvement ideas

Get **gyroscope** sensor data to cluster unique motion

No arm involvement,
long duration



Similar repetition,
long duration

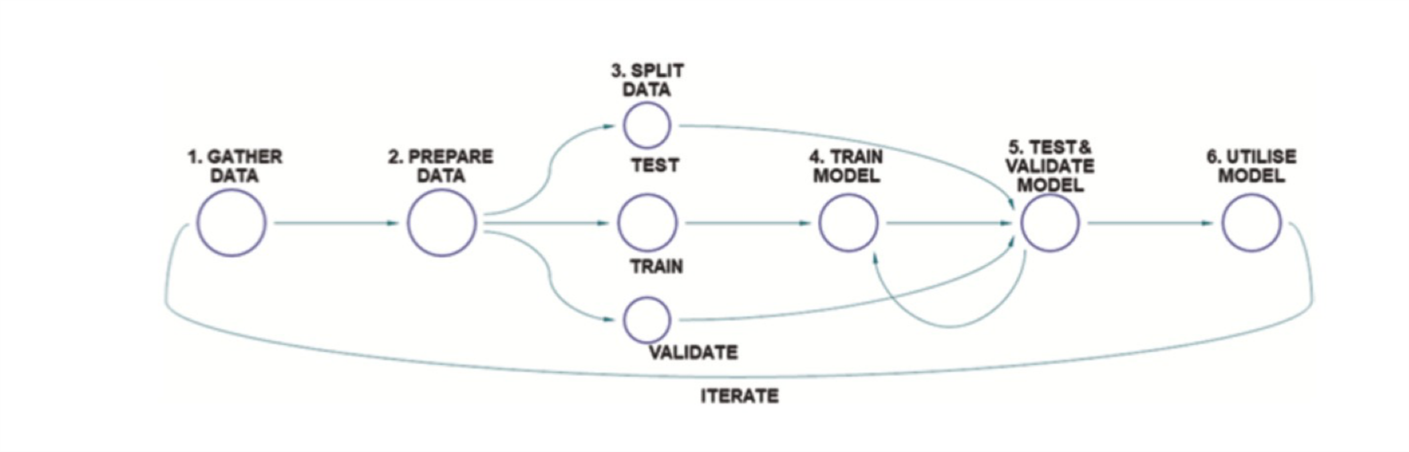


Similar repetition,
short duration

4. K-Nearest Neighbors (KNN)
Limitations

- Training data vs. test data
  - But: Real world implications
- Accuracy not majorly influenced by k
- Might average speed have been enough?

5. Data Workflow and Challenges
Machine Learning Workflow



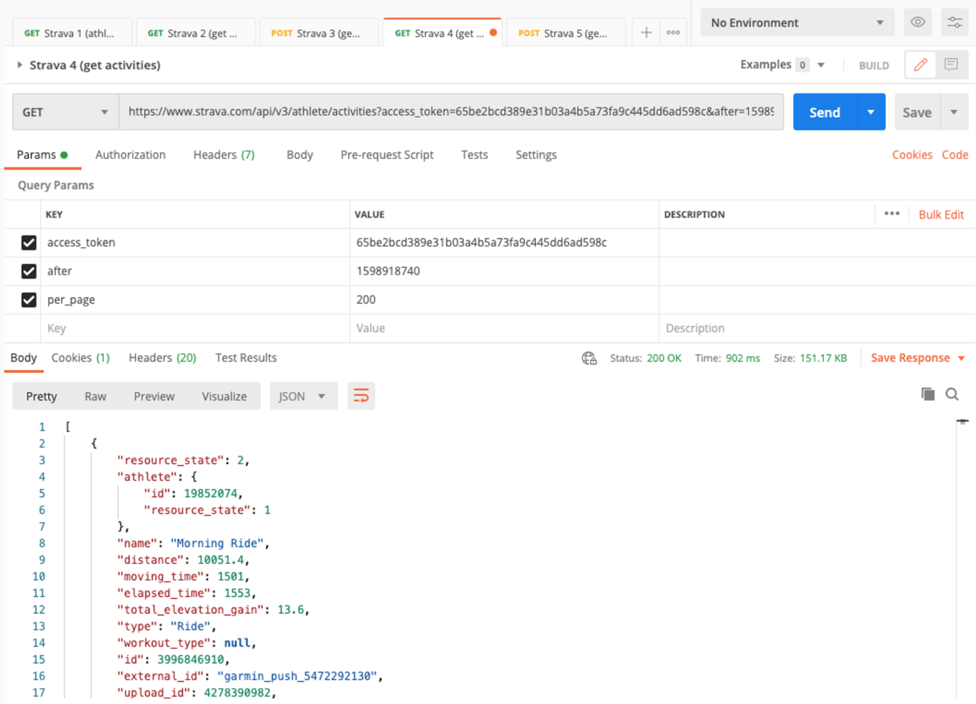
Yang, 2020

5. Data Workflow and Challenges
1. Gather Data

STRAVA™

POSTMAN

CSVJSON
Sponsored by Flatfile.io



5. Data Workflow and Challenges

2. Prepare Data

- Add country/city names from latitude/longitude for each GPS-recorded activity with Python (reverse-geocode 1.4.1)

```
>>> import reverse_geocode
>>> coordinates = (-37.81, 144.96), (31.76, 35.21)
>>> reverse_geocode.search(coordinates)
[{'city': 'Melbourne', 'code': 'AU', 'country': 'Australia'},
 {'city': 'Jerusalem', 'code': 'IL', 'country': 'Israel'}]
```

- Manipulate values to make more readable
  - Distance from meters to kilometers (/ 1000)
  - Moving time from seconds to (/ 3600)
  - Average speed from meters per second to kilometers per hour (* 3.6)

IB3K50

Artificial Intelligence for Business

Gaspard Dangy

Philipp John

Hugo Jouan

Martin Knaze

Virgile Siac

Léna-Lou Simon--Virolle

Roberta Sildmäe

Johannes Pittgens

December 14, 2020

Task 1 : References

Cohen, A. (2020) Get Smart: AI And The Energy Sector Revolution. Forbes [online] Available at: <https://www.forbes.com/sites/arielcohen/2020/08/31/get-smart-ai-and-the-energy-sector-revolution/?sh=4dc60d0c6044> [Accessed: 13 December 2020]

Chen, G. (2017) "Leveraging smart data for predictive maintenance in wind industry." Vestas. [Online] Available at: https://www.asiacleanenergyforum.org/wp-content/uploads/2017/06/2_leveraging-smart-data-for-predictive-maintenance-in-wind-industry.pdfn [Accessed: 28 November 2020]

Dedezade, E., (2019), *Jobs of the future; how self piloted AI drones are creating exciting opportunities*, Microsoft (online) https://news.microsoft.com/europe/features/jobs-of-the-future-how-self-piloted-ai-drones-are-creating-exciting-new-opportunities/ [Accessed: 13 December 2020]

Deloitte (2015) "The future of the global power sector : Preparing for emerging opportunities and threats" [Online] Available at: https://www2.deloitte.com/content/dam/Deloitte/tr/Documents/energy-resources/power-future-global-power-sector-report.pdf [Accessed: 28 November 2020]

Task 1 : References

Deloitte (2019) 2020 power and utilities industry outlook. [online] Available at: <https://www2.deloitte.com/xe/en/pages/energy-and-resources/articles/power-and-utilities-industry-outlook.html> (Accessed: 13 December 2020)

Dezide, (2020) Vestas captures global knowledge in intelligent software from Dezide (online) Available at: https://www.dezide.com/customers/vestas-captures-global-knowledge-experience-dezide/ [Accessed: 14 December 2020]

Forsthoffer, WE. (2011) *Forsthoffer's Best Practice Handbook For Rotating Machinery,* Butterworth-Heinemann.

Hashemian, H. M. (2011) "State-of-the-Art Predictive Maintenance Techniques," in IEEE Transactions on Instrumentation and Measurement, vol. 60, no. 1, pp. 226-236, Jan. 2011, doi: 10.1109/TIM.2010.2047662. [Accessed: 28 November 2020]

IBM (n.d.). *Maximo APM - Predictive Maintenance Insights*. [online] www.ibm.com. Available at: https://www.ibm.com/products/ibm-maximo-asset-performance-management/predictive-maintenance-insights?lnk=STW_US_STESCH&lnk2=learn_MaxPredInsight&pexp=def&psrc=none&mhsrc=ibmsearch_a&mhq=predictive%20maintenance [Accessed 12 Dec. 2020].

Task 1 : References

IEA (2020), Renewables 2020, IEA, Paris https://www.iea.org/reports/renewables-2020

IRENA (2019), Future of wind: Deployment, investment, technology, grid integration and socio-economic aspects (A Global Energy Transformation paper), International Renewable Energy Agency, Abu Dhabi.

Kent, R. (2018). *Energy Management in Plastics Processing : Strategies, Targets, Techniques, and Tools*. San Diego, Ca, Usa: Elsevier Science.

McKinsey Global Institute, (2014) Artificial Intelligence, the next digital frontier?, McKinsey (online). https://www.mckinsey.com/~/media/mckinsey/industries/advanced%20electronics/our%20insights/how%20artificial%20intelligence%20can%20deliver%20real%20value%20to%20companies/mgi-artificial-intelligence-discussion-paper.ashx [Accessed: 11 December 2020]

Miller, T.W. (2014). *Modeling techniques in predictive analystics with Python and R : a guide to data science.* Upper Saddle River, Nj: Pearson Education.

Mobley, K. (2001). *Plant Engineer's Handbook*. Butterworth-Heinemann, pp.312–317.

Task 1 : References

Newman, N. (2019) “AI: the energy industry’s untapped resource” in Engineering & Technology, Volume 14, Issue 7, Pages 66-69, ISSN 1750-9645, [Online] Available at: https://eandt.theiet.org/content/articles/2019/07/ai-the-energy-industry-s-untapped-resource/
[Accessed: 26 November 2020]

Next (n.d.) What is Artificial Intelligence in the Energy Industry? [online] Available at: <https://www.next-kraftwerke.com/knowledge/artificial-intelligence> [Accessed: 13 December 2020]

Quinonero-Candela, J., Ido Dagan, D’alché-Buc, F. and Magnini, B. (2006). *Machine Learning Challenges : Evaluating Predictive Uncertainty, Visual Object Classification, and Recognizing Textual Entailment, First Pascal Machine Learning Challenges Workshop, MLCW 2005, Southampton, UK, April 11-13.* Berlin ; Heidelberg: Springer.

Schlechtingen, M. and Ferreira Santos, I. (2018) Comparative analysis of neural network and regression based condition monitoring approaches for wind turbine fault detection. *Mechanical Systems and Signal Processing*. 25(4), pp. 1849-1875. [Accessed: 28 November 2020]

Stetco, A. et al. (2019) Machine learning methods for wind turbine condition monitoring: A review. *Renewable Energy*. 133(7), pp. 620-635. [Accessed: 24 November 2020]

Task 1 : References

Suganthi L., Anand A. Samuel, (2012) “Energy models for demand forecasting—A review”, in renewable and Sustainable Energy Reviews, Volume 16, Issue 2, 2012, Pages 1223-1240, ISSN 1364-0321, https://doi.org/10.1016/j.rser.2011.08.014. [Online] Available at: http://www.sciencedirect.com/science/article/pii/S1364032111004242 [Accessed: 29 November 2020]

Yoo, C.-H.; Chung, I.-Y.; Lee, H.-J.; Hong, S.-S. (2013) “ Intelligent Control of Battery Energy Storage for Multi-Agent Based Microgrid Energy Management. Energies”, 2013, 6, 4956-4979. [Accessed: 30 November 2020]

Wang, KS., Sharma, V.S. & Zhang, ZY. SCADA data based condition monitoring of wind turbines. Adv. Manuf. 2, 61–69 (2014). https://doi.org/10.1007/s40436-014-0067-0

Task 2 - References

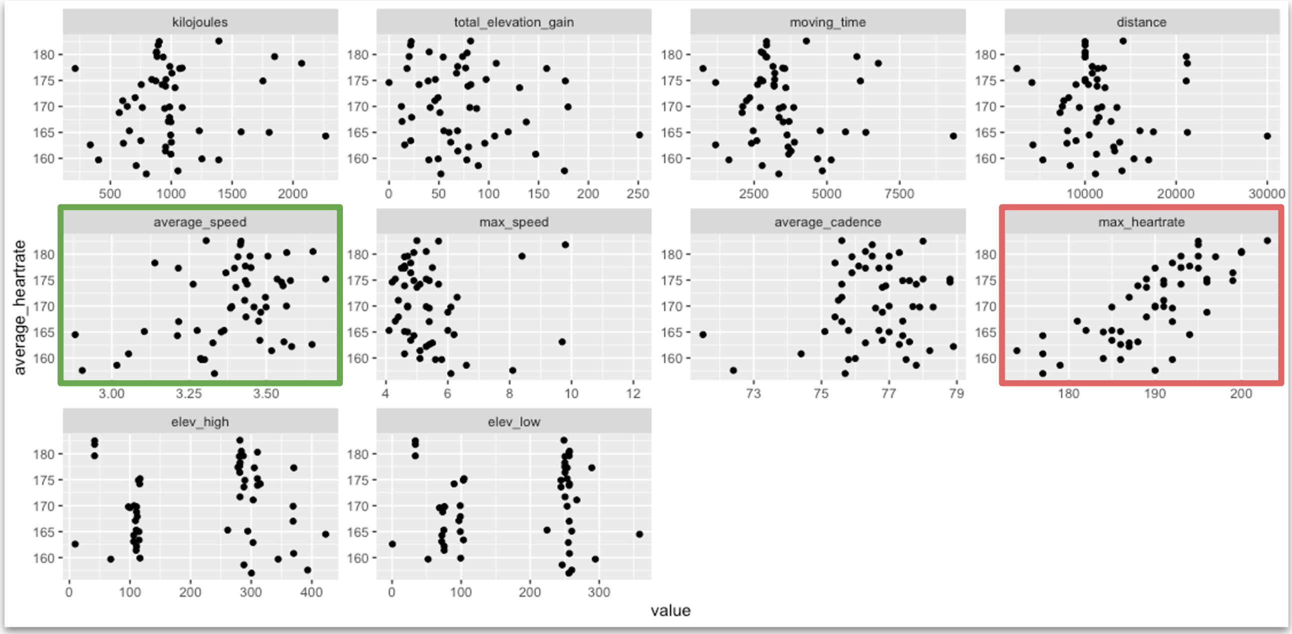
Lambert, M.I., Mbambo, Z.H. and Gibson, A.S.C. (1998). Heart rate during training and competition for long distance running. *Journal of Sports Sciences*, 16(sup1), pp.85–90.

Padulo, J., Powell, D., Milia, R. and Ardigò, L.P. (2013). A Paradigm of Uphill Running. *PLoS ONE*, [online] 8(7). Available at: https://doi.org/10.1371/journal.pone.0069006 [Accessed 4 Dec. 2020].

Yang, Y. (2020) *Lecture 5: Machine Learning I.* Lecture Slides. IB3K50 Artificial Intelligence for Business, Warwick Business School.

Appendix

Appendix 1: Relationships between each independent variable and the average heart rate



Findings

- No clear relationships between most independent variables and avg. heart rate
- Avg. speed has a sight positive correlation and is independent of heart rate
- Max. heart rate is obviously positively correlated with avg. heart rate, but not independent of it and can therefore not be used. Furthermore, this data was not collected in the given scenario.

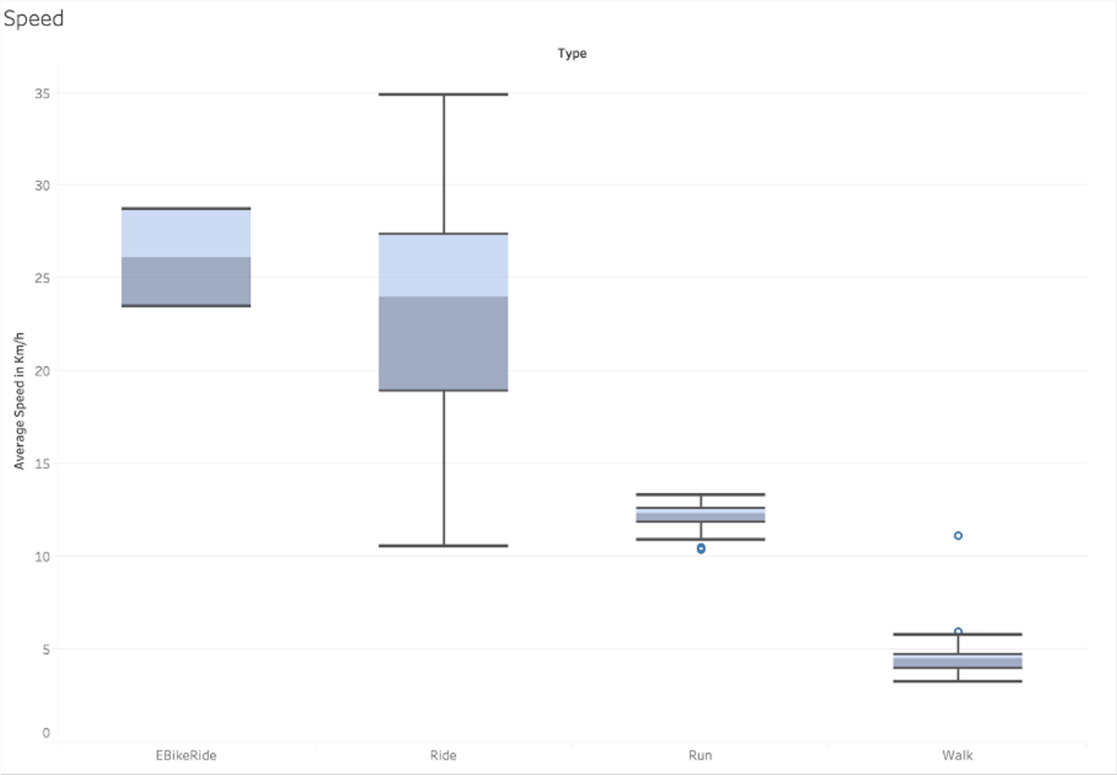
Appendix 2: Pearson's correlation table to check for multicollinearity

	total_elevation_gain	moving_time	distance
total_elevation_gain	1.0000000	0.3147616	0.2486019
moving_time	0.3147616	1.0000000	0.9911176
distance	0.2486019	0.9911176	1.0000000

Findings from Pearson's correlation table

- No multicollinearity between moving time and total elevation gain
- Strong positive correlation between distance and moving time (dependent variable)
- Weak positive correlation between total elevation gain and moving time

Appendix 3: Box and whisker plot for Average Speed

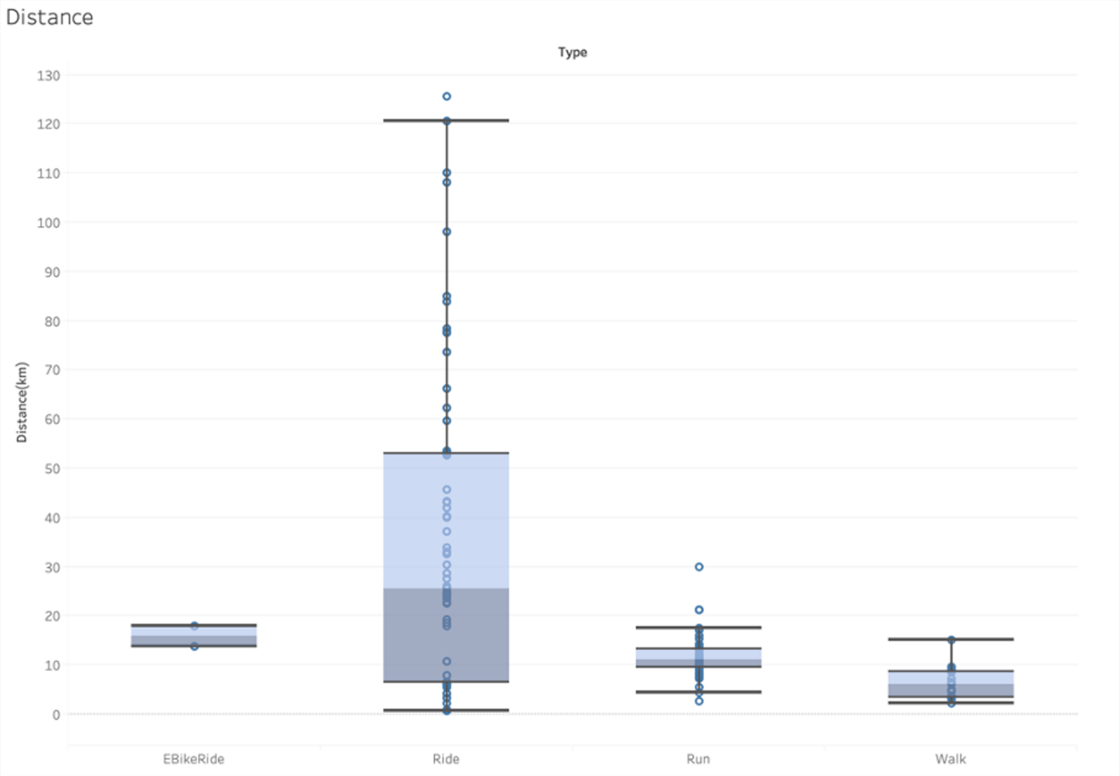


Maximum Value
- Run: 13 km/h
- Ride: 34 km/h

Range
- Run: 3 km/h
- Ride: 24 km/h

Median
- Run: 12 km/h
- Ride: 24 km/h

Appendix 3: Box and whisker plot for Distance

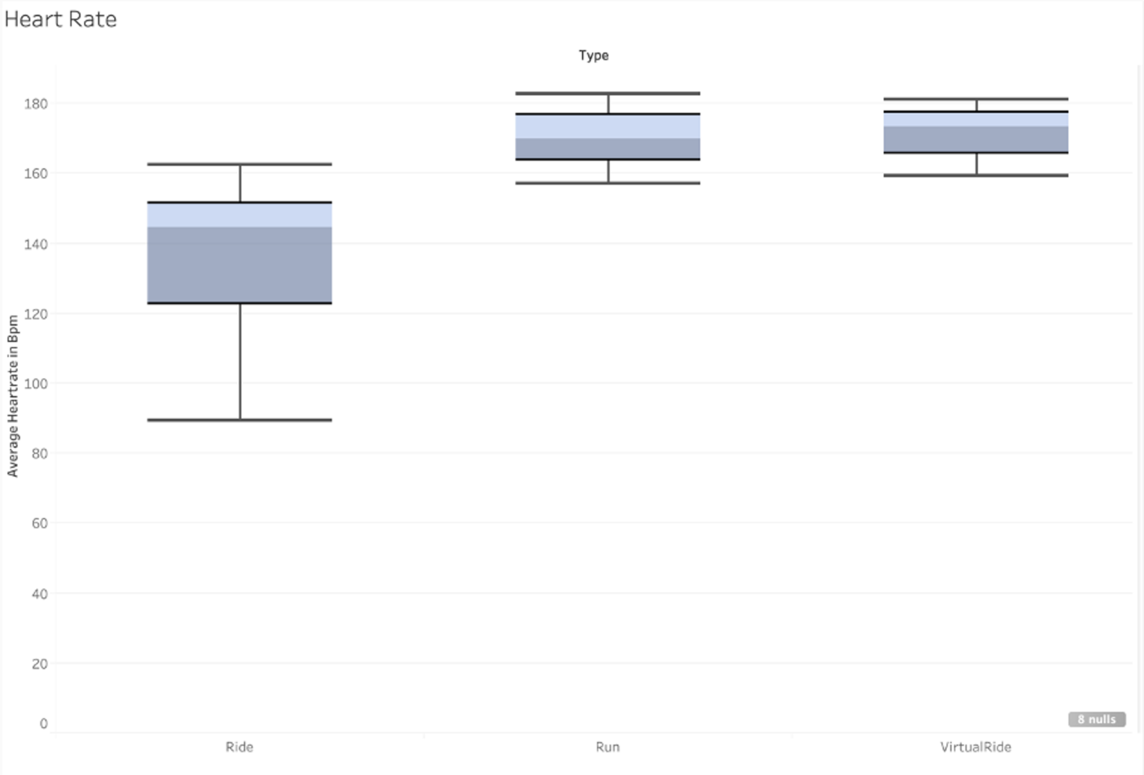


Maximum Value
- Run: 17 km
- Ride: 120 km

Range
- Run: 13 km
- Ride: 100 km

Median
- Run: 11 km
- Ride: 25 km

Appendix 3: Box and whisker plot for Average Heart Rate



Maximum Value
- Run: 182 bpm
- Ride: 162 bpm

Range
- Run: 25 bpm
- Ride: 73 bpm

Median
- Run: 169 bpm
- Ride: 144 bpm