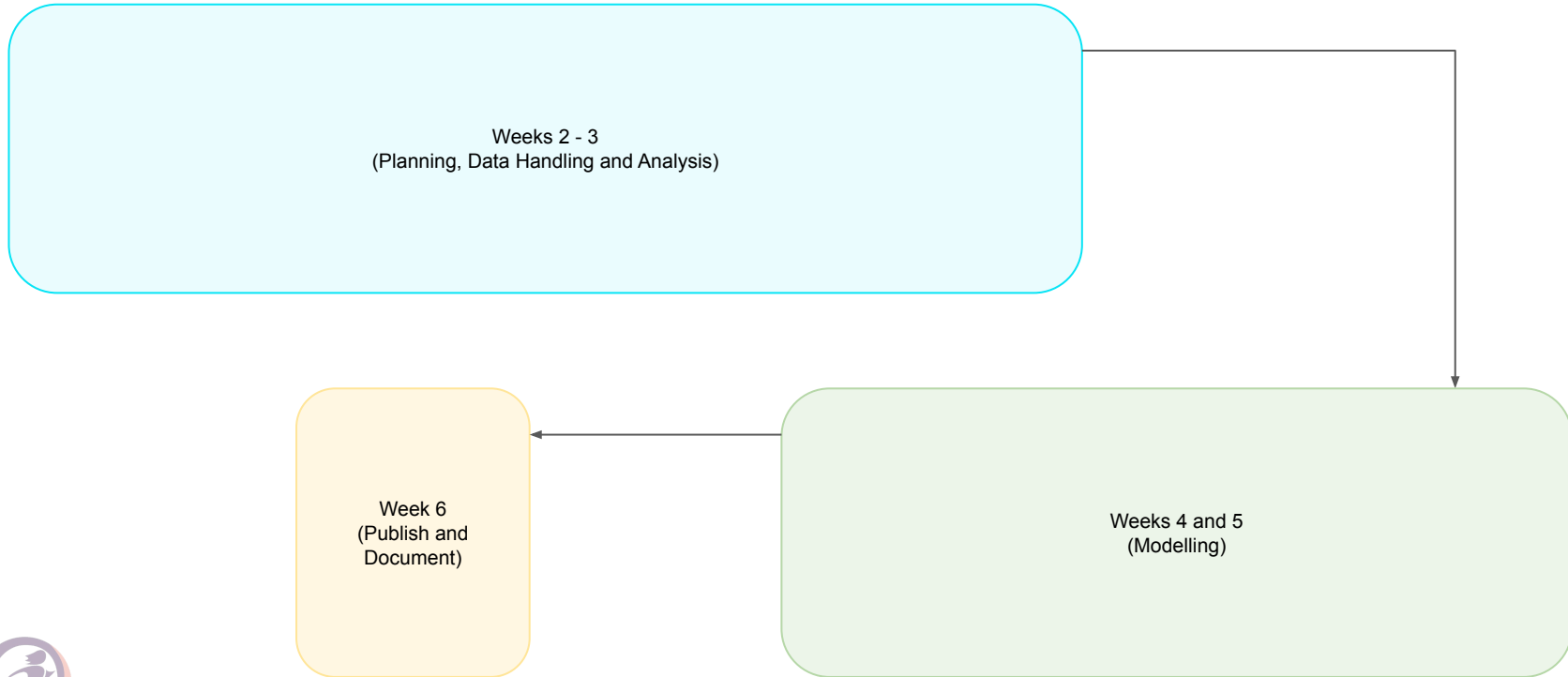


Week 4 panel session

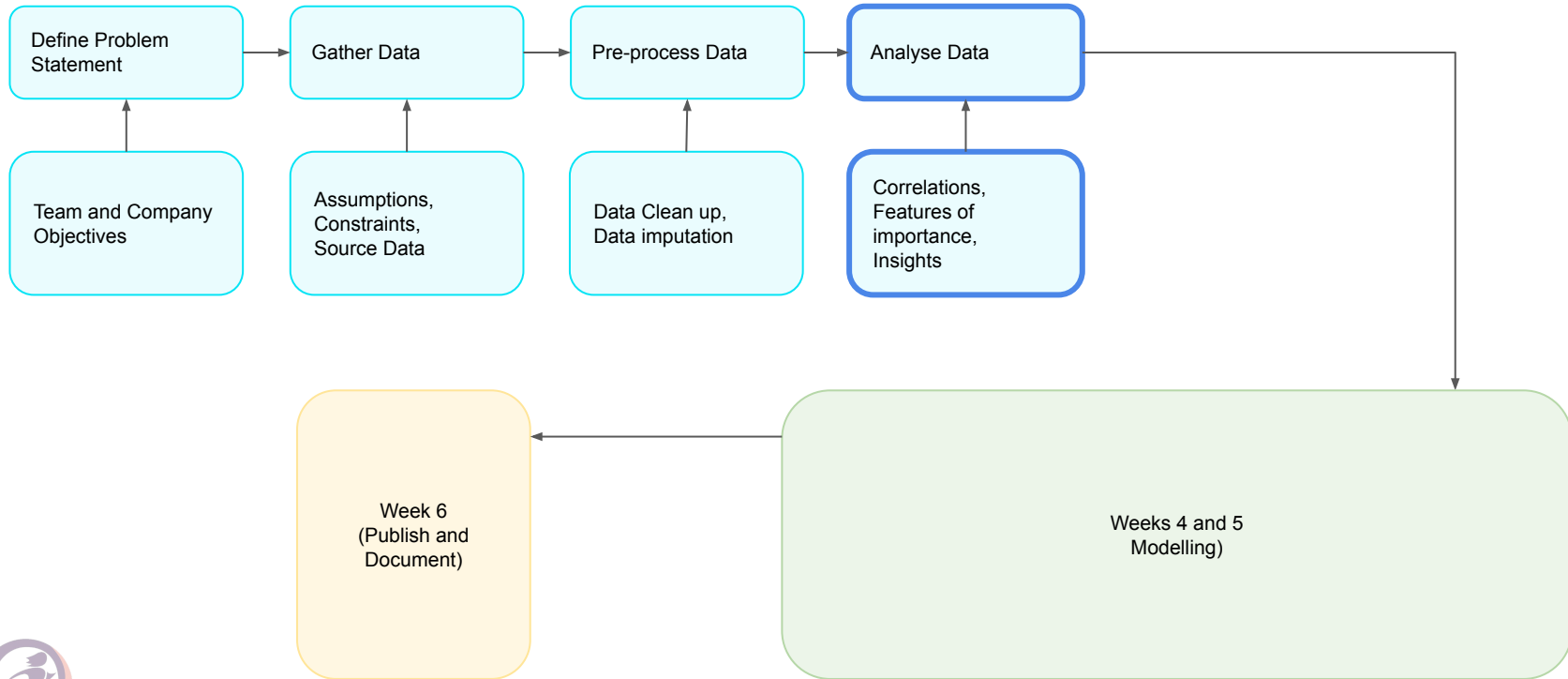
Project 5: Workout Analysis Modeling and Data Visualisation

Redback Operations - Data Science and Analytics

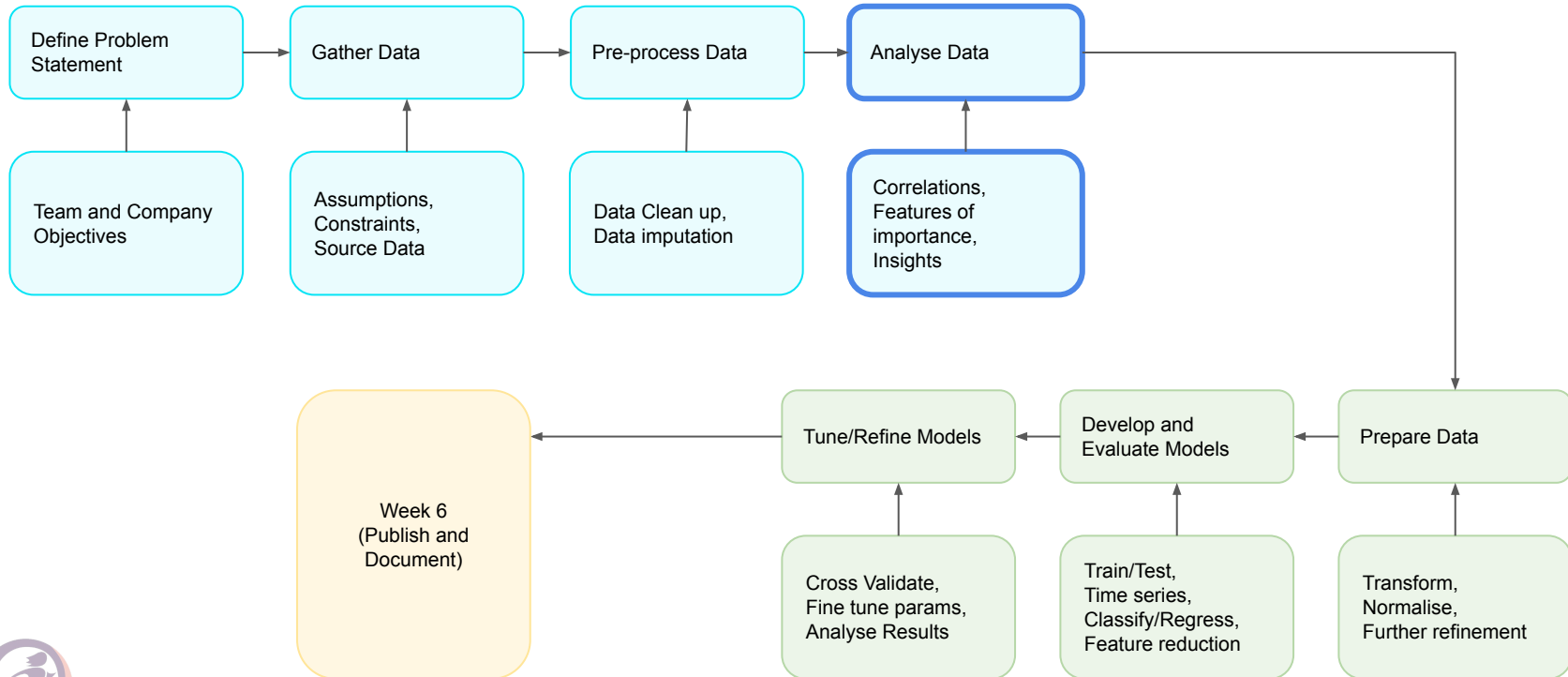
Approach: Machine Learning (ML) Workflow



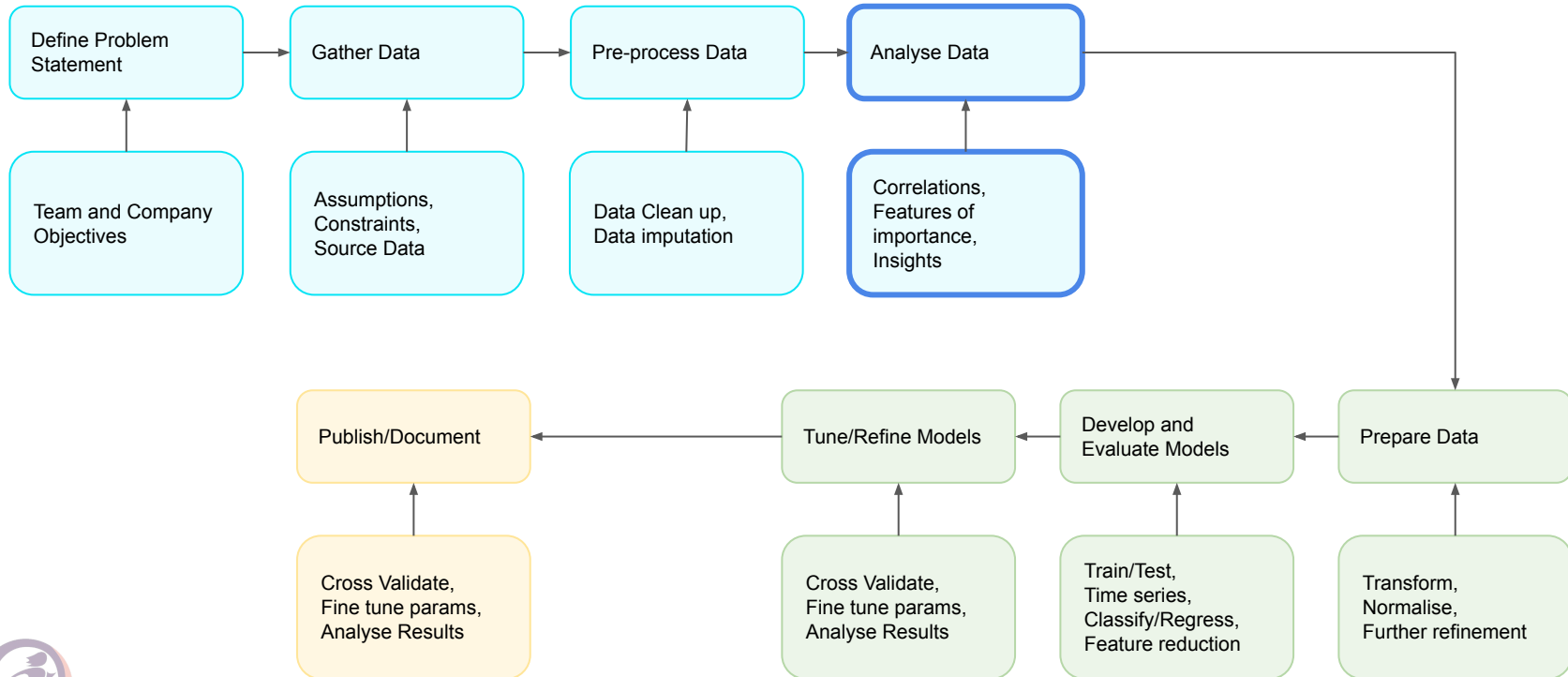
Approach: Machine Learning (ML) Workflow



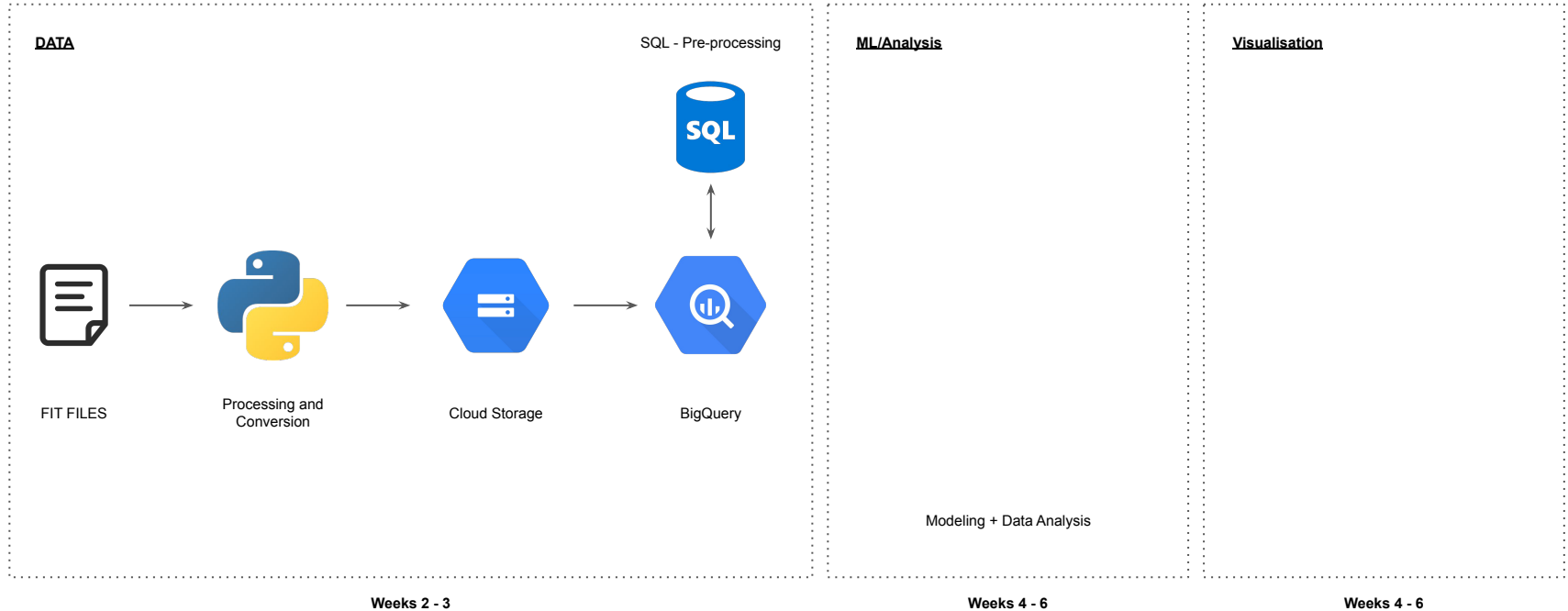
Approach: Machine Learning (ML) Workflow



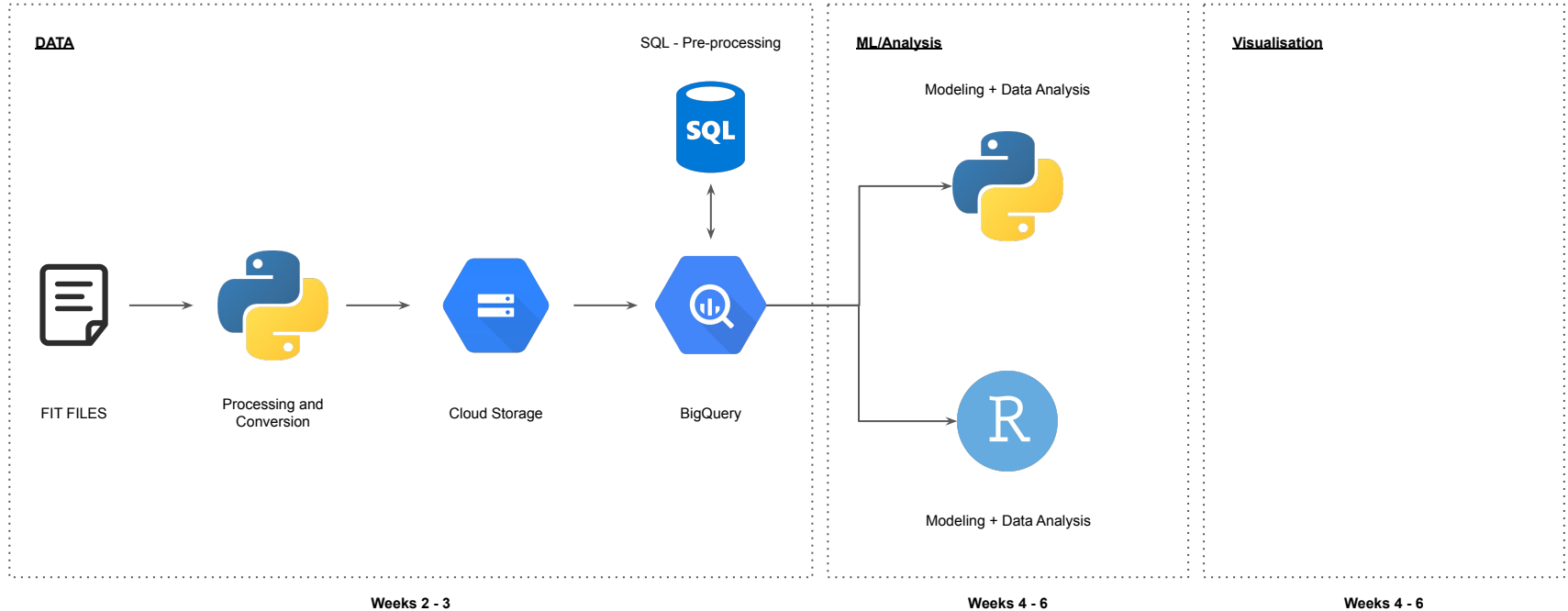
Approach: Machine Learning (ML) Workflow



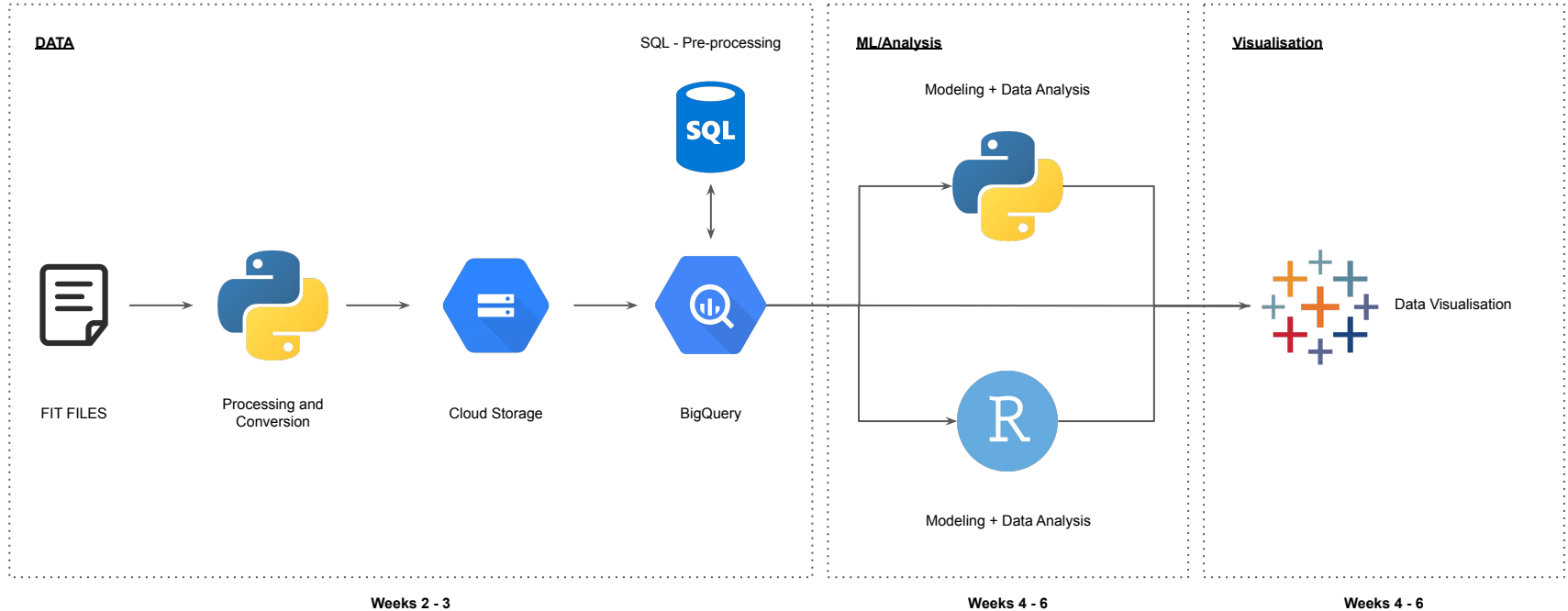
Tech Stack



Tech Stack



Tech Stack



Progress

Conversion Script

The screenshot displays a Python IDE with a project named 'FITConversion.py'. The left sidebar shows a file explorer with a list of files: 'activity-file-10.fit', 'activity-file-20.fit', 'activity-file-30.fit', 'activity-file-40.fit', 'activity-file-50.fit', 'activity-file-60.fit', 'activity-file-70.fit', 'FITConversion', 'activity-file-10', 'activity-file-20', 'activity-file-30', 'activity-file-40', 'activity-file-50', 'activity-file-60', 'activity-file-70', and 'combined_csv'. The main editor shows the script 'FITConversion.py' with the following code:

```
1 # MTELLEY DEAKIN
2 # SIT174
3 # DATA SCIENCE AND ANALYSIS
4
5 import ...
6
7 allowed_fields = [...]
8 required_fields = ["timestamp"]
9
10 UTC = pytz.UTC
11 AEST = pytz.timezone("Australia/Melbourne")
12
13 def renameFiles():...
14
15 def main():...
16
17 def write_fitfile_to_csv(fitfile, output_file="test_output.csv"):...
18
19 def combine_csvfiles():...
20
21 if __name__ == "__main__":
22     renameFiles()
23     main()
24     combine_csvfiles()
25
26 ...
```

The bottom panel shows the 'Run' output for 'FITConversion (1)'. The output text is:

```
Renaming Started
C:\Users\markt\Wahoo_FIT_Conversion\Export\RUN
7 files renamed
Conversion started:
File Converted
File Converted
File Converted
File Converted
File Converted
File Converted
File Converted
File Converted
finished conversions
Combining Files
Process finished with exit code 0
```

Two pink arrows point from the 'Script Execution' and 'Output' labels to the 'Run' output panel. The 'Script Execution' arrow points to the 'Renaming Started' line, and the 'Output' arrow points to the '7 files renamed' line.

Progress

Table Schema

```
{ schema_wahoo_V2.json /Users/marktelley/Documents/GitHub/data-analysis/Trimester 3 2022/Project 5 Workout Analysis and Model/Dataset Procure } x ...  
1  [  
2    {  
3      "name": "timestamp",  
4      "mode": "NULLABLE",  
5      "type": "STRING",  
6      "description": "time UTC",  
7      "fields": []  
8    },  
9    {  
10     "name": "position_lat",  
11     "mode": "NULLABLE",  
12     "type": "STRING",  
13     "fields": []  
14   },  
15   {  
16     "name": "position_long",  
17     "mode": "NULLABLE",  
18     "type": "STRING",  
19     "fields": []  
20   },  
21   {  
22     "name": "distance",  
23     "mode": "NULLABLE",  
24     "type": "STRING",  
25     "description": "km"  
26   },  
27   {  
28     "name": "enhanced_altitude",  
29     "mode": "NULLABLE",  
30     "type": "STRING",  
31     "description": "m",  
32     "fields": []  
33   },  
34   {  
35     "name": "altitude",
```

Progress

SQL

```
cleanse_script_V2.sql /Users/marktelley/Documents/GitHub/data-analysis/Trimester 3 2022/Project 5 Workout Analysis and Model/Dataset Procurement/SQL/cleanse_script_V2.sql

1  -- Author: MTELLEY
2  -- Release 1
3  -- CLEANSE AND ORGANISE DATA FROM CSV FILE
4  -- TO DO: Create SessionID using rank/row ops (Version 2)
5  -- //////////////////////////////////
6  SELECT
7  -- WORKOUT METRICS
8  timestamp,
9  TIMESTAMP(DATETIME(CAST(CONCAT(SPLIT(timestamp, '+')[SAFE_ORDINAL(1)], " ", "UTC") AS TIMESTAMP), "Australia/Melbourne")) AS timestamp,
10 CAST((TIMESTAMP(DATETIME(CAST(CONCAT(SPLIT(timestamp, '+')[SAFE_ORDINAL(1)], " ", "UTC") AS TIMESTAMP), "Australia/Melbourne"))) AS DATETIME) AS timestamp,
11 -- position_lat, -- removed for privacy reasons
12 -- position_long, -- removed for privacy reasons
13 SAFE_CAST(distance AS FLOAT64) AS distance,
14 SAFE_CAST(enhanced_altitude AS FLOAT64) AS enhanced_altitude,
15 -- altitude, -- removed
16 SAFE_CAST(ascent AS FLOAT64) AS ascent,
17 SAFE_CAST(grade AS FLOAT64) AS grade,
18 SAFE_CAST(calories AS FLOAT64) AS calories,
19 SAFE_CAST(enhanced_speed AS FLOAT64) AS enhanced_speed,
20 -- speed, -- removed
21 SAFE_CAST(heart_rate AS FLOAT64) AS heart_rate,
22 SAFE_CAST(temperature AS INT64) AS temperature,
23 SAFE_CAST(cadence AS FLOAT64) AS cadence,
24 SAFE_CAST(power AS FLOAT64) AS power,
25 ROUND(SAFE_CAST(left_right_balance AS FLOAT64)/100,2) AS left_right_balance,
26 SAFE_CAST(gps_accuracy AS FLOAT64) AS gps_accuracy,
27 -- PRODUCT DETAILS:
28 -- descriptor,
29 -- product_name,
30 -- serial_number, -- removed for privacy reasons
31 sessionID,
32 -- TO DO
33 CASE
34 | WHEN Timestamp IS NOT NULL THEN "U1000000" -- USER ID, MANUALLY SET
35 ELSE
36 NULL
37 END
38 AS userID,
39 CASE
40 | WHEN Timestamp IS NOT NULL THEN 33 -- AGE, MANUALLY SET
41 ELSE
42 NULL
43 END
44 AS age,
45 CASE
46 | WHEN Timestamp IS NOT NULL THEN "MALE" -- GENDER, MANUALLY SET
47 ELSE
48 NULL
49 END
50 AS gender.
```

Progress

SQL

masterjoin.sql /Users/marktelley/Documents/GitHub/data-analysis/Trimester 3 2022/Project 5

```
1  -- MASTER FILE OPERATIONS
2  SELECT
3      timestamp,
4      timestamp_AEST,
5      date_AEST,
6      distance,
7      enhanced_altitude,
8      ascent,
9      grade,
10     calories,
11     enhanced_speed,
12     heart_rate,
13     temperature,
14     cadence,
15     power,
16     -- left_right_balance -- REMOVED DUE TO ERRORS (Garmin VS Wahoo related issue)
17     gps_accuracy,
18     sessionID,
19     userID,
20     age,
21     gender,
22     weight,
23     FTP,
24 FROM
25     `redbackoperationsdataai.Fitness_Data.fitness-activity-*` -- WILDCARD
26
```

```
Q Master_Ops x +
RUN SAVE SHARE SCHEDULE MORE
1  -- MASTER FILE OPERATIONS
2  SELECT
3      userID,
4      ROUND(AVG(heart_rate),2) AS avg_heart_rate,
5      ROUND(AVG(enhanced_speed),2) AS avg_speed,
6      ROUND(AVG(power),2) AS avg_power,
7      ROUND(AVG(cadence),2) AS avg_cadence,
8      MAX(age) AS age,
9      MAX(FTP) AS FTP,
10     MAX(weight) AS weight,
11     trim(gender) as gender,
12     ROUND(MAX(FTP)/MAX(weight),2) AS wperkg,
13 FROM
14     `redbackoperationsdataai.Fitness_Data.fitness-activity-*`
15 WHERE
16     power IS NOT NULL AND heart_rate IS NOT NULL
17 GROUP BY
18     userID,
19     gender
```

Query results

JOB INFORMATION		RESULTS	JSON	EXECUTION DETAILS		EXECUTION GRAPH		PREVIEW		
Row	userID	avg_heart_rate	avg_speed	avg_power	avg_cadence	age	FTP	weight	gender	wperkg
1	U1000002	163.52	27.5	184.82	78.17	46	270	75	MALE	3.6
2	U1000000	155.75	30.26	204.79	74.06	33	301	80	MALE	3.76
3	U1000004	151.93	25.56	168.5	64.64	34	280	74	MALE	3.78
4	U1000010	129.63	26.3	135.95	66.54	64	298	86	MALE	3.47
5	U1000008	157.92	27.6	152.75	76.33	16	250	64	MALE	3.91
6	U1000003	162.23	30.73	233.4	74.81	33	310	81	MALE	3.83
7	U1000009	144.15	26.64	203.62	71.09	46	335	118	MALE	2.84

Progress

Google Cloud

The screenshot shows the Google Cloud Storage 'Bucket details' page for 'redback_operations_bucket'. The bucket is located in 'us' and has a 'Standard' storage class. The 'OBJECTS' tab is selected, showing a list of files. A filter is applied: 'Filter by name prefix only' with the filter text 'ExerciseData'. The list shows several CSV files named 'combined_csv_user1.csv' through 'combined_csv_user9.csv', all created in November 2022.

Name	Size	Type	Created	Storage class	Last
Archv/	—	Folder	—	—	—
combined_csv_user1.csv	256.3 MB	text/csv	Nov 27, 2022, 1:09:19 PM	Standard	Nov
combined_csv_user10.csv	596.3 MB	text/csv	Nov 27, 2022, 7:52:17 PM	Standard	Nov
combined_csv_user2.csv	6.3 MB	text/csv	Nov 27, 2022, 1:08:34 PM	Standard	Nov
combined_csv_user3.csv	43 MB	text/csv	Nov 27, 2022, 1:09:47 PM	Standard	Nov
combined_csv_user4.csv	118.9 MB	text/csv	Nov 27, 2022, 5:40:02 PM	Standard	Nov
combined_csv_user5.csv	503 MB	text/csv	Nov 27, 2022, 7:35:18 PM	Standard	Nov
combined_csv_user6.csv	248.3 MB	text/csv	Nov 27, 2022, 5:50:07 PM	Standard	Nov
combined_csv_user7.csv	254.5 MB	text/csv	Nov 27, 2022, 5:58:34 PM	Standard	Nov
combined_csv_user8.csv	288.5 MB	text/csv	Nov 27, 2022, 7:25:07 PM	Standard	Nov
combined_csv_user9.csv	212.2 MB	text/csv	Nov 27, 2022, 6:07:02 PM	Standard	Nov

The screenshot shows the Google Cloud BigQuery interface. The 'redbackoperationsdataai' dataset is selected in the Explorer. The 'master-fitness-activity' table is highlighted. The 'DETAILS' tab is active, showing table information and storage details. A red box highlights the 'Number of rows' value, 15,740,893, which is used in the calculation below.

Table ID	redbackoperationsdataai.Master_Fitness_Data.master-fitness-activity
Created	Nov 27, 2022, 1:55:14 PM UTC+11
Last modified	Nov 27, 2022, 8:44:59 PM UTC+11
Table expiration	NEVER
Data location	US
Default collation	
Description	

Storage info	
Number of rows	15,740,893
Total logical bytes	2.08 GB
Active logical bytes	2.08 GB
Long term logical bytes	0 B
Total physical bytes	215.17 MB
Active physical bytes	215.17 MB
Long term physical bytes	0 B
Time travel physical bytes	67.28 MB

$$(15.74 \times 10^6) / 3600 = 4375 \text{ hours}$$

Progress

Documentation

Redback Operations
Data Science and Analytics Team
M.TELLEY



DOCUMENTATION *Data Procurement, Cleansing and Organisation*

Purpose	1
Scope	2
Wahoo and Garmin fitness activity data procurement	2
What is / Who are Garmin and Wahoo?	2
File output - FIT files	2
Where was the data procured from?	3
Accessing the data	3
Other information collected	3
Privacy	3
FIT File Conversion, Data cleansing and Organisation	4
FIT Conversion	4
Data Cleansing and Organisation	4
Key Operations (Wahoo)	5
Key Operations (Garmin)	13
Access to BigQuery	13
Outcome	13

Purpose

Redback Operation core product will collect performance (output) data from a number of sensors while the user is using the bike and wahoo kickr trainer. Moreover, establishing a working environment that will align with the core product is critical in establishing continuity in data-related tasks and advising other core teams about Data Science and Analytics (DSA) Team's requirements. The initial effort is to set up a core data location for the team that house all key datasets.

By the end of this task, further work on data analysis, visualisation and using a number of machine learning algorithms will be possible.

MTELLEY

1

Progress

Features

features_cleanse

feature_refinement

RUN

MORE

SAVE

SHARE

SCHEDULE

```
1 CREATE TEMP FUNCTION
2 maxheart_rate(x FLOAT64)
3 RETURNS FLOAT64 AS ( 211-0.64*x );
4 CREATE TEMP FUNCTION
5 timeofday(x INT64)
6 RETURNS STRING AS(
7 CASE
8 WHEN x >= 4 AND x < 12 THEN "MORNING"
9 WHEN x >= 12 AND x < 18 THEN "AFTERNOON"
10 WHEN (x < 4 AND x > 8) OR x >= 18 THEN "EVENING"
11 ELSE
12 "N/A"
13 END
14 );
15 CREATE TEMP FUNCTION
16 weekend(x DATE)
17 RETURNS BOOL AS (
18 IF
19 (EXTRACT(DAYOFWEEK
20 FROM
21 x) = 1
22 OR EXTRACT(DAYOFWEEK
23 FROM
24 x) = 7, TRUE, FALSE));
25 CREATE TEMP FUNCTION
26 durationalc(distance FLOAT64,
27 speed FLOAT64)
28 RETURNS INT64 AS( CAST((distance*3600)/speed AS INT64) );
29 WITH
30 user AS (
31 SELECT
32 *,
33 power =
34 IF
```

Query results

JOB INFORMATION		RESULTS	JSON	EXECUTION DETAILS		EXECUTION GRAPH		PREVIEW									
Row	userID	date_AEST	distance	enhanced_avg_s	avg_power	avg_wpkg	avg_heart_rate	avg_max_heart	enhanced_durat	weeken	day_of_week	hour	5	day_cat			
9	U1000000	2021-04-04	22.58	27.92	204.25	2.55	150.58	0.79	00:48:31	true	1	5	5	MORNING			
10	U1000000	2021-04-06	40.01	30.49	251.56	3.14	127.0	0.67	01:18:45	false	3	5	5	MORNING			
11	U1000000	2021-04-17	77.02	26.52	238.47	2.98	151.08	0.8	02:54:15	true	7	6	6	MORNING			
12	U1000000	2021-04-24	75.69	26.63	232.46	2.91	152.0	0.8	02:50:32	true	7	6	6	MORNING			
13	U1000000	2021-05-01	31.09	30.72	232.41	2.91	121.36	0.64	01:00:44	true	7	6	6	MORNING			
14	U1000000	2021-05-07	65.11	30.24	234.01	2.93	106.45	0.56	02:09:11	false	6	14	14	AFTERNOON			
15	U1000000	2021-05-16	37.11	27.63	202.61	2.53	125.67	0.66	01:20:36	true	1	9	9	MORNING			
16	U1000000	2021-05-21	46.91	31.14	202.9	2.54	147.03	0.77	01:30:23	false	6	6	6	MORNING			
17	U1000000	2021-05-27	41.33	30.07	231.86	2.9	145.17	0.76	01:22:28	false	5	16	16	AFTERNOON			
18	U1000000	2021-05-29	31.38	31.27	258.25	3.23	141.98	0.75	01:00:13	true	7	10	10	MORNING			
19	U1000000	2021-06-02	44.51	29.03	228.32	2.85	104.44	0.55	01:32:00	false	4	8	8	MORNING			
20	U1000000	2021-06-04	25.66	27.84	254.54	3.18	106.88	0.56	00:55:18	false	6	16	16	AFTERNOON			
21	U1000000	2021-06-05	37.89	30.78	241.83	3.02	110.11	0.58	01:13:51	true	7	8	8	MORNING			
22	U1000000	2021-06-11	25.37	29.6	229.27	2.87	118.47	0.62	00:51:26	false	6	16	16	AFTERNOON			
23	U1000000	2021-06-13	105.02	32.49	234.9	2.94	112.05	0.59	03:13:56	true	1	5	5	MORNING			
24	U1000000	2021-06-14	26.0	23.63	209.47	2.62	115.5	0.61	01:06:01	false	2	12	12	AFTERNOON			
25	U1000000	2021-06-15	40.32	26.72	214.09	2.68	99.8	0.53	01:30:32	false	3	5	5	MORNING			
26	U1000000	2021-06-17	64.96	28.96	209.91	2.62	102.41	0.54	02:14:35	false	5	5	5	MORNING			
27	U1000000	2021-06-19	61.31	25.35	263.96	3.3	117.49	0.62	02:25:06	true	7	7	7	MORNING			
28	U1000000	2021-06-20	118.52	30.52	206.09	2.58	100.97	0.53	03:53:00	true	1	6	6	MORNING			
29	U1000000	2021-06-22	40.33	29.99	260.64	3.26	120.72	0.64	01:20:41	false	3	5	5	MORNING			

feature-summary-filtered

QUERY

SHARE

SCHEMA

DETAILS

PREVIEW

Filter

Enter property name or value

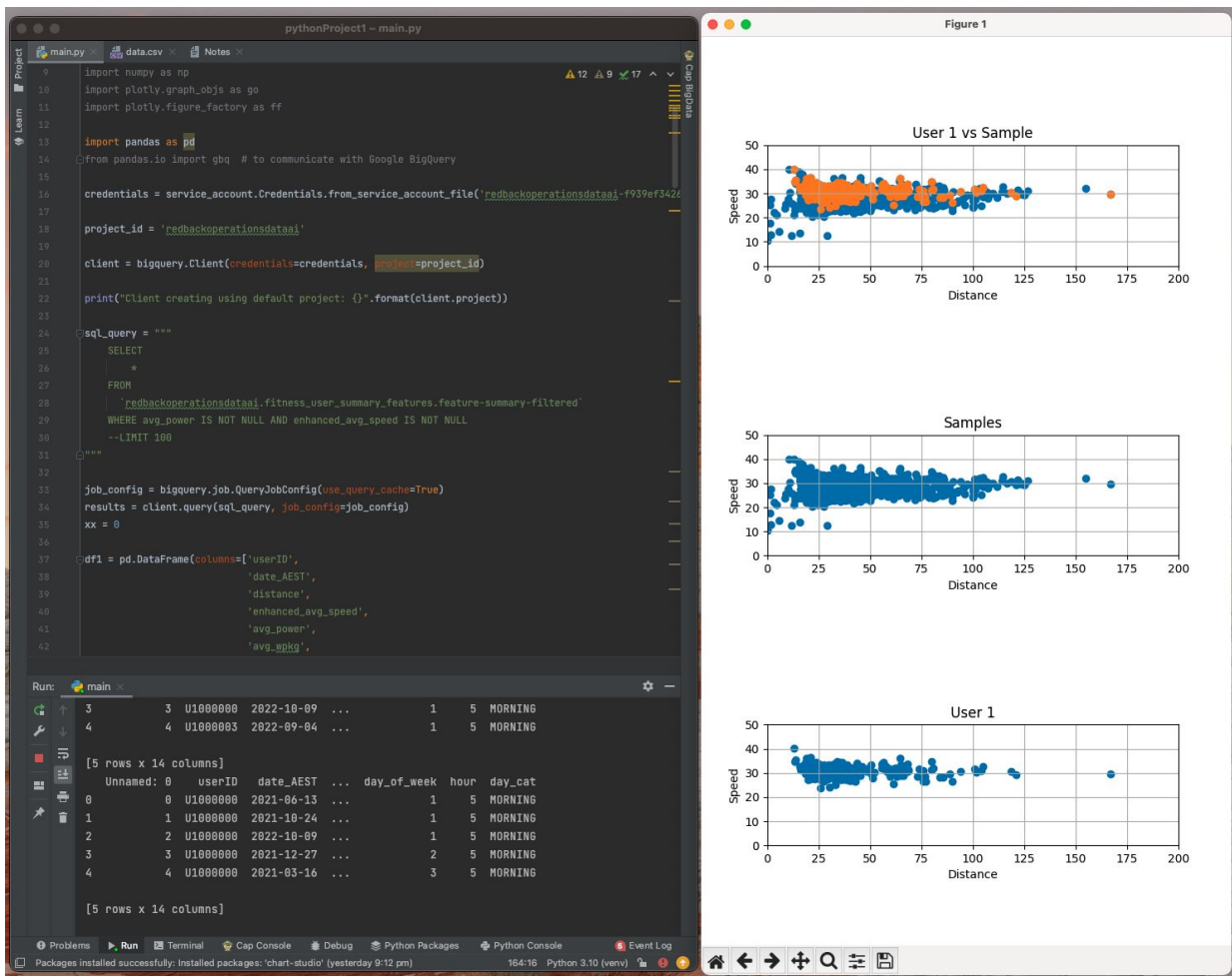
	Field name	Type	Mode	Collation
<input type="checkbox"/>	userID	STRING	NULLABLE	
<input type="checkbox"/>	date_AEST	DATE	NULLABLE	
<input type="checkbox"/>	distance	FLOAT	NULLABLE	
<input type="checkbox"/>	enhanced_avg_speed	FLOAT	NULLABLE	
<input type="checkbox"/>	avg_power	FLOAT	NULLABLE	
<input type="checkbox"/>	avg_wpkg	FLOAT	NULLABLE	
<input type="checkbox"/>	avg_heart_rate	FLOAT	NULLABLE	
<input type="checkbox"/>	avg_max_heart_rate_perct	FLOAT	NULLABLE	
<input type="checkbox"/>	enhanced_duration	TIME	NULLABLE	
<input type="checkbox"/>	weekend	BOOLEAN	NULLABLE	
<input type="checkbox"/>	day_of_week	INTEGER	NULLABLE	
<input type="checkbox"/>	hour	INTEGER	NULLABLE	
<input type="checkbox"/>	day_cat	STRING	NULLABLE	



Progress

Bigquery API and Python

Starting to prep for analysis & insights



Next

Connecting either RStudio or Pycharm with Bigquery to create a data stream.

Research Time Series models; predicting HR/Power etc

Develop and Evaluate Models using either R or Python

Using .py or .r scripts within Tableau to create custom data views.

Thank you