

Technical Task

TEKEVWE KWAKPOVWE

Data Cleaning

- Change column names to make it easier to write code with

Data Cleaning

- Change column names to make it easier to write code with
 - 'Pitch_x' → 'x'
 - 'Pitch_y' → 'y'
 - 'participation_id' → 'id'
 - 'Time (s)' → 'time'
 - 'Speed (m/s)' → 'speed'

Data Cleaning

- Change column names to make it easier to write code with
 - 'Pitch_x' → 'x'
 - 'Pitch_y' → 'y'
 - 'participation_id' → 'id'
 - 'Time (s)' → 'time'
 - 'Speed (m/s)' → 'speed'
- Filter values to only include positions within valid regions of the pitch

Data Cleaning

- Change column names to make it easier to write code with
 - 'Pitch_x' → 'x'
 - 'Pitch_y' → 'y'
 - 'participation_id' → 'id'
 - 'Time (s)' → 'time'
 - 'Speed (m/s)' → 'speed'
- Filter values to only include positions within valid regions of the pitch
- Sort the values for more efficient grouping

Data Cleaning

- Change column names to make it easier to write code with
 - 'Pitch_x' → 'x'
 - 'Pitch_y' → 'y'
 - 'participation_id' → 'id'
 - 'Time (s)' → 'time'
 - 'Speed (m/s)' → 'speed'
- Filter values to only include positions within valid regions of the pitch
- Sort the values for more efficient grouping
 - By 'id' and 'time'

Data Cleaning

- Change column names to make it easier to write code with
 - 'Pitch_x' → 'x'
 - 'Pitch_y' → 'y'
 - 'participation_id' → 'id'
 - 'Time (s)' → 'time'
 - 'Speed (m/s)' → 'speed'
- Filter values to only include positions within valid regions of the pitch
- Sort the values for more efficient grouping
 - By 'id' and 'time'
- Map IDs to new values for interpretability

Leaderboard Generation

Filter →

Leaderboard Generation

Filter → Groupby 'id'→

Leaderboard Generation

Filter → Groupby 'id' → Function Call →

Leaderboard Generation

Filter → Groupby 'id' → Function Call → Output

Leaderboard Generation

Filter → Groupby 'id' → Function Call → Output

```
lboard1 = df[df['id'] != 'ball'][['id', 'x', 'y']]
```

```
lboard2 = df[(df['id'] != 'ball') &  
             (df['speed'] > 5.5) &  
             (df['speed'] < 6.97)][['id', 'x', 'y']]
```

```
lboard3 = df[df['id'] != 'ball'][['id', 'speed']]
```

Leaderboard Generation

Filter → Groupby 'id' → Function Call → Output

```
lboard1 = df[df['id'] != 'ball'][['id', 'x', 'y']]
```

→

```
.apply(Lambda x: np.sqrt((x['x'].diff()*2) +
(x['y'].diff()*2)).sum(),
include_groups=False)
```

```
lboard2 = df[(df['id'] != 'ball') &
(df['speed'] > 5.5) &
(df['speed'] < 6.97)][['id', 'x', 'y']]
```

→

```
lboard3 = df[df['id'] != 'ball'][['id', 'speed']]
```

→

```
max_speed = lboard3.groupby('id').max().sort
```

Leaderboards

Total Distance

id	total_distance
Kyra	11997.300936
Alessia	11517.437233
Lia	11450.617449
Frida	11015.236440
Kim	9969.641081
Mariona	9709.671493
Chloe	8536.513427
Daphne	8217.648173
Katie	7745.246281
Stina	4276.831954
Steph	3115.694337
Emily	1598.497534
Leah	991.271944
Laia	982.629093

Distance Zone 2

	id	distance_zone2
0	Frida	2461.970263
1	Kyra	2359.354563
2	Kim	1423.309597
3	Alessia	1216.034475
4	Mariona	941.911000
5	Chloe	913.869905
6	Daphne	805.991163
7	Lia	719.581942
8	Katie	706.193814
9	Steph	548.841893
10	Stina	449.473451
11	Emily	125.372260
12	Laia	65.194673

Max Speed

id	speed
Frida	8.200
Laia	7.789
Mariona	7.693
Kim	7.683
Steph	7.559
Daphne	7.528
Kyra	7.136
Chloe	7.076
Stina	6.997
Alessia	6.990
Katie	6.969
Lia	6.724
Emily	6.644
Leah	3.693

Heatmap

Method

Heatmap

Method

- Round up positions

Heatmap

Method

- Round up positions
- Place in bins

Heatmap

Method

- Round up positions
- Place in bins:
 - 60 categories for ‘x’
 - 40 categories for ‘y’

Heatmap

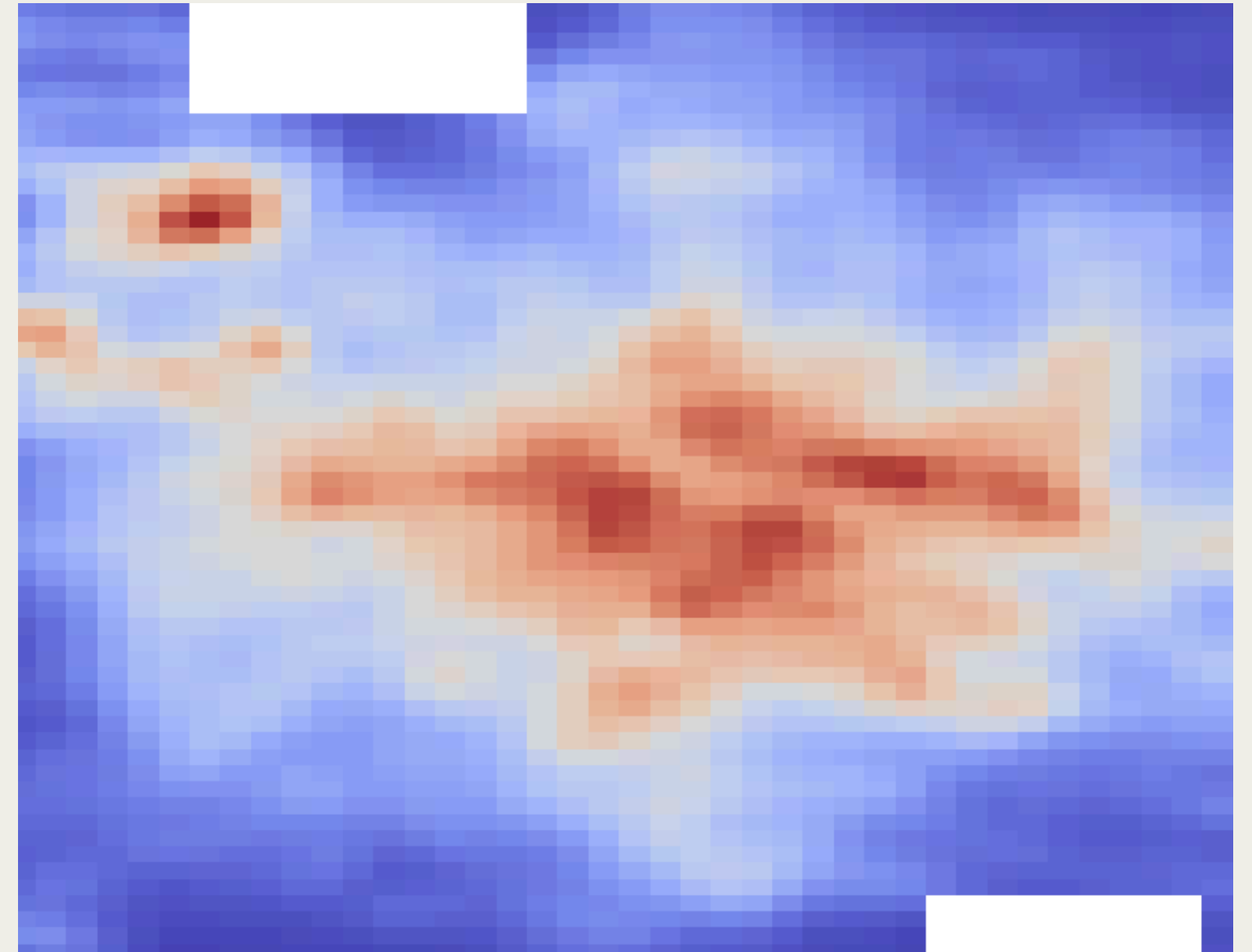
Method

- Round up positions
- Place in bins:
 - 60 categories for 'x'
 - 40 categories for 'y'
- Smooth using Gaussian filter

Heatmap

Method

- Round up positions
- Place in bins:
 - 60 categories for 'x'
 - 40 categories for 'y'
- Smooth using Gaussian filter



Ball Integration

How much possession did each player have?

Ball Integration

How much possession did each player have?

- possession → proximity to ball

Ball Integration

How much possession did each player have?

- possession → proximity to ball
- Establish some threshold

Ball Integration

How much possession did each player have?

- possession → proximity to ball
- Establish some threshold
- Accumulate time

Ball Integration

How much possession did each player have?

- possession → proximity to ball
- Establish some threshold
- Accumulate time

```
ball spent approximately 75.8 seconds with Kim  
ball spent approximately 45.0 seconds with Kyra  
ball spent approximately 40.3 seconds with Alessia  
ball spent approximately 36.8 seconds with Frida  
ball spent approximately 21.2 seconds with Lia  
ball spent approximately 20.4 seconds with Chloe  
ball spent approximately 16.1 seconds with Katie  
ball spent approximately 12.2 seconds with Steph  
ball spent approximately 8.7 seconds with Mariona  
ball spent approximately 3.5 seconds with Stina  
ball spent approximately 2.7 seconds with Daphne  
ball spent approximately 0.6 seconds with Leah
```

Some Observations

- **The ball was out of play a lot**
- **Heat map suggests 1 goal**
- **Positions**
 - Kyra, Alessia and Frida are midfielders
 - Laia was the keeper