



Fundamental Concepts in Data Insight:

Demo: Automating Insight

Fundamentals for a General Audience





QA Ltd. owns the copyright and other intellectual property rights of this material and asserts its moral rights as the author. All rights reserved.

```
import pandas as pd
```





The Simulation

Suppose we're trying to predict the risk of victimization.

One method here is to keep a table of risk factors that can be applied to any individual person. The *weights* of these factors can be determined from historical datasets and even expert judgment.

```
risk_factors
```

```
name      0.0  
arrests   0.5  
age       0.2  
dtype: float64
```





We multiply each of these factors by what we observe a person to have,

```
alice
```

```
name      Alice
arrests    10
age        18
dtype: object
```

```
(
    risk_factors["arrests"] * alice["arrests"] +
    risk_factors["age"]      * alice["age"]
) / 2
```

4.3





We can generalise this to table of people,

```
pd.DataFrame?
```

```
people
```

	arrests	age
name		
Alice	10	18
Bob	10	21
Eve	10	35
Lucie	10	35
Alex	10	35

```
results = (people * factors).mean(1)
```

```
results
```

```
name
Alice    4.3
Bob      4.6
Eve      6.0
Lucie    6.0
Alex     6.0
dtype: float64
```





Descriptive Analytics

These are the type of metrics we would include in a report.

Highest risk person,

```
results.idxmax(), results.max()
```

```
('Eve', 6.0)
```





Lowest risk person,

```
results.idxmin(), results.min()
```

```
('Alice', 4.3)
```

Median risk,

```
results.median()
```

```
6.0
```

People with the median,

```
results[ results == results.median() ]
```

```
name
Eve      6.0
Lucie    6.0
Alex     6.0
dtype: float64
```

A sample of people,





How do you automate insight?

When building an automation system we will often want to make a **decision** based on these type of measures,

```
risk_threshold = 5

for name, risk in results.items():
    if risk > risk_threshold:
        print(f"ALERT: {name} above threshold!")
```

```
ALERT: Eve above threshold!
ALERT: Lucie above threshold!
ALERT: Alex above threshold!
```





Exercise

- Revise the table of people, add in additional observables
- Update the factors to include a risk weight for each factor
- Revise these weights until the risk totals make sense

