

# Harnessing AI in Financial Services: Laying the Data Foundations

This notebook demonstrates how data quality and governance directly affect the outcomes of AI models, using a simplified example of customer data.

We'll simulate loan approval predictions based on customer age and income, showing how clean, accurate data improves model performance.

```
In [45]: import pandas as pd
import numpy as np
from sklearn.linear_model import LogisticRegression
from sklearn.inputs import SimpleInputer
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
from pydicom import baseModel, field_validator
from typing import Optional
import warnings
warnings.filterwarnings("ignore")
```

## 1. Raw customer data

Albeit well-structured, the data is inaccurate with non-sensical ages, incomes, deposits, and mortgage amounts. All credit values are stored as 600, and loan approvals are also incorrectly recorded.

```
In [46]: # Example raw data from a finance application process
bad_data = pd.read_csv(r"C:\Users\GabriellaDixon\Projects\AI Thought Leadership\synthetic_financial_data_bad_quality.csv")
bad_data
```

|    | Age   | Annual_Income | Home_Value | Deposit   | Mortgage_Amount | Credit_Score | Loan_Approved |
|----|-------|---------------|------------|-----------|-----------------|--------------|---------------|
| 0  | 999.0 | -0.00         | 518668.13  | 63351.39  | 350969.93       | 600          | 1             |
| 1  | 42.0  | 31929.64      | 309610.67  | 128519.52 | 157719.96       | 600          | 0             |
| 2  | 60.0  | 68894.28      | 503270.83  | 28744.17  | 461823.14       | 600          | 1             |
| 3  | 55.0  | 125827.35     | 205746.32  | 22165.76  | 340733.18       | 600          | 0             |
| 4  | 32.0  | 41804.75      | 226363.80  | 94199.03  | 284147.34       | 600          | 0             |
| 5  | 48.0  | 136386.27     | 177935.41  | 20012.84  | 335973.36       | 600          | 1             |
| 6  | 57.0  | 131538.83     | 406812.65  | 22276.14  | 171915.56       | 600          | 0             |
| 7  | 67.0  | 125874.02     | 275941.69  | 33830.76  | 91247.29        | 600          | 0             |
| 8  | 72.0  | 0.00          | 535456.14  | 40519.95  | 443045.50       | 600          | 0             |
| 9  | 58.0  | 45535.24      | 414954.05  | 25541.78  | 419888.64       | 600          | 0             |
| 10 | 104.0 | 52949.46      | 227990.52  | 82190.48  | 281732.50       | 600          | 1             |
| 11 | 43.0  | 148899.29     | 534550.56  | 51070.26  | 232304.44       | 600          | 0             |
| 12 | 35.0  | 43727.59      | 257598.42  | 29916.40  | 446942.45       | 600          | 0             |
| 13 | 32.0  | 922577.00     | 5562955.20 | 80885.09  | 385037.65       | 600          | 0             |
| 14 | 75.0  | 44851.13      | 563632.03  | 49366.86  | 420686.27       | 600          | 0             |
| 15 | 71.0  | 39478.49      | 116789.07  | 9624.87   | 353938.30       | 600          | 0             |
| 16 | 42.0  | 0.00          | 169759.17  | 36477.82  | 208545.29       | 600          | 0             |
| 17 | 50.0  | 828429.60     | 5373742.50 | 16499.70  | 248363.53       | 600          | 0             |
| 18 | 29.0  | 99519.46      | 272824.39  | 10400.33  | 199954.19       | 600          | 0             |
| 19 | 57.0  | 1057891.00    | 4282416.15 | 42329.35  | 136106.50       | 600          | 0             |
| 20 | 0.0   | 114491.79     | 498300.96  | 86716.34  | 196874.15       | 600          | 0             |
| 21 | 73.0  | 146567.82     | 238010.49  | 116215.47 | 373284.43       | 600          | 1             |
| 22 | 56.0  | 77127.57      | 331923.23  | 29512.36  | 409411.89       | 600          | 1             |
| 23 | 63.0  | 125245.42     | 539496.33  | 64952.52  | 450198.05       | 600          | 1             |
| 24 | 62.0  | 0.00          | 179497.64  | 67246.35  | 144706.76       | 600          | 0             |
| 25 | 25.0  | 2377211.80    | 7272617.40 | 37634.80  | 125702.54       | 600          | 1             |
| 26 | 73.0  | 1211709.80    | 6691609.35 | 47737.39  | 398369.90       | 600          | 0             |
| 27 | 38.0  | 32132.11      | 430578.31  | 40676.73  | 122899.78       | 600          | 0             |
| 28 | 64.0  | 135877.16     | 303586.29  | 18031.32  | 290642.70       | 600          | 0             |
| 29 | 74.0  | 96756.84      | 459123.91  | 10089.12  | 451421.78       | 600          | 0             |
| 30 | 999.0 | 1667140.20    | 4924162.50 | 129986.40 | 138908.85       | 600          | 0             |
| 31 | 75.0  | 96636.58      | 350042.12  | 66268.60  | 149073.36       | 600          | 0             |
| 32 | 53.0  | -588040.00    | 2199875.85 | 96450.83  | 169097.31       | 600          | 1             |
| 33 | 50.0  | 76392.32      | 140728.32  | 83559.70  | 225653.72       | 600          | 0             |
| 34 | 69.0  | 31680.82      | 305074.01  | 31918.84  | 178623.66       | 600          | 0             |
| 35 | 28.0  | 136508.14     | 172165.47  | 15025.78  | 364924.88       | 600          | 0             |
| 36 | 45.0  | 121028.74     | 154795.88  | 87608.11  | 203303.93       | 600          | 0             |
| 37 | 66.0  | 46792.58      | 243685.04  | 88486.42  | 114243.15       | 600          | 1             |
| 38 | 42.0  | 118225.14     | 151742.77  | 126044.25 | 443659.85       | 600          | 0             |
| 39 | 58.0  | 1562240.00    | 2185970.55 | 52182.88  | 149191.24       | 600          | 1             |
| 40 | -1.0  | 0.00          | 438146.92  | 91301.74  | 246259.28       | 600          | 0             |
| 41 | 55.0  | 36403.12      | 519518.57  | 97413.74  | 339276.57       | 600          | 0             |
| 42 | 37.0  | 114648.65     | 346068.52  | 37499.62  | 197333.76       | 600          | 0             |
| 43 | 42.0  | 124725.13     | 171239.12  | 75858.72  | 354629.77       | 600          | 0             |
| 44 | 64.0  | 148447.25     | 263142.75  | 17818.51  | 320529.76       | 600          | 1             |
| 45 | 18.0  | 2206978.00    | 7231753.35 | 23758.61  | 206528.16       | 600          | 1             |
| 46 | 54.0  | 122363.89     | 161238.20  | 79947.29  | 155665.77       | 600          | 0             |
| 47 | 36.0  | 131402.32     | 433885.59  | 23341.51  | 283573.45       | 600          | 0             |
| 48 | 54.0  | -20.00        | 7637590.35 | 75987.50  | 265183.43       | 600          | 0             |
| 49 | 49.0  | 109528.01     | 479489.45  | 12039.21  | 447416.56       | 600          | 0             |

## 2. Split the data for training and testing

```
In [47]: bad_x = bad_data[['Age', 'Annual_Income', 'Home_Value', 'Deposit', 'Mortgage_Amount', 'Credit_Score']]
bad_y = bad_data['Loan_Approved']
bad_x_train, bad_x_test, bad_y_train, bad_y_test = train_test_split(bad_x, bad_y, random_state=42)
```

```
In [48]: bad_x_train
```

|    | Age   | Annual_Income | Home_Value | Deposit   | Mortgage_Amount | Credit_Score |
|----|-------|---------------|------------|-----------|-----------------|--------------|
| 8  | 72.0  | 0.00          | 535456.14  | 40519.95  | 443045.50       | 600          |
| 3  | 55.0  | 125827.35     | 205746.32  | 22165.76  | 340733.18       | 600          |
| 6  | 57.0  | 131538.83     | 406812.65  | 22276.14  | 171915.56       | 600          |
| 41 | 55.0  | 36403.12      | 519518.57  | 97413.74  | 339276.57       | 600          |
| 46 | 54.0  | 122363.89     | 161238.20  | 79947.29  | 155665.77       | 600          |
| 47 | 36.0  | 131402.32     | 433885.59  | 23341.51  | 283573.45       | 600          |
| 15 | 71.0  | 39478.49      | 116789.07  | 9624.87   | 353938.30       | 600          |
| 9  | 58.0  | 45535.24      | 414954.05  | 25541.78  | 419888.64       | 600          |
| 16 | 42.0  | 0.00          | 169759.17  | 36477.82  | 208545.29       | 600          |
| 24 | 62.0  | 0.00          | 179497.64  | 67246.35  | 144706.76       | 600          |
| 34 | 69.0  | 31680.82      | 305074.01  | 31918.84  | 178623.66       | 600          |
| 31 | 75.0  | 96636.58      | 350042.12  | 66268.60  | 149073.36       | 600          |
| 0  | 999.0 | -0.00         | 518668.13  | 63351.39  | 350969.93       | 600          |
| 44 | 64.0  | 148447.25     | 263142.75  | 17818.51  | 320529.76       | 600          |
| 27 | 38.0  | 32132.11      | 430578.31  | 40676.73  | 122899.78       | 600          |
| 33 | 50.0  | 76392.32      | 140728.32  | 83559.70  | 225653.72       | 600          |
| 5  | 48.0  | 136386.27     | 177935.41  | 20012.84  | 335973.36       | 600          |
| 29 | 74.0  | 96756.84      | 459123.91  | 10089.12  | 451421.78       | 600          |
| 11 | 43.0  | 148899.29     | 534550.56  | 51070.26  | 232304.44       | 600          |
| 36 | 45.0  | 121028.74     | 154795.88  | 87608.11  | 203303.93       | 600          |
| 1  | 42.0  | 31929.64      | 309610.67  | 128519.52 | 157719.96       | 600          |
| 21 | 73.0  | 146567.82     | 238010.49  | 116215.47 | 373284.43       | 600          |
| 2  | 60.0  | 68894.28      | 503270.83  | 28744.17  | 461823.14       | 600          |
| 43 | 42.0  | 124725.13     | 171239.12  | 75858.72  | 354629.77       | 600          |
| 35 | 28.0  | 136508.14     | 172165.47  | 15025.78  | 364924.88       | 600          |
| 23 | 63.0  | 125245.42     | 539496.33  | 64952.52  | 450198.05       | 600          |
| 40 | -1.0  | 0.00          | 438146.92  | 91301.74  | 246259.28       | 600          |
| 10 | 104.0 | 52949.46      | 227990.52  | 82190.48  | 281732.50       | 600          |
| 22 | 56.0  | 77127.57      | 331923.23  | 29512.36  | 409411.89       | 600          |
| 18 | 29.0  | 99519.46      | 272824.39  | 10400.33  | 199954.19       | 600          |
| 49 | 49.0  | 109528.01     | 479489.45  | 12039.21  | 447416.56       | 600          |
| 20 | 0.0   | 114491.79     | 498300.96  | 86716.34  | 196874.15       | 600          |
| 7  | 67.0  | 125874.02     | 275941.69  | 33830.76  | 91247.29        | 600          |
| 42 | 37.0  | 114648.65     | 346068.52  | 37499.62  | 197333.76       | 600          |
| 14 | 75.0  | 44851.13      | 563632.03  | 49366.86  | 420686.27       | 600          |
| 28 | 64.0  | 135877.16     | 303586.29  | 18031.32  | 290642.70       | 600          |
| 38 | 42.0  | 118225.14     | 151742.77  | 126044.25 | 443659.85       | 600          |

## 3. Train a simple model on the data

```
In [49]: bad_model = LogisticRegression()
bad_model.fit(bad_x_train, bad_y_train)
bad_preds = bad_model.predict(bad_x_test)
```

```
print("Model Accuracy:", accuracy_score(bad_y_test, bad_preds))
Model Accuracy: 0.46153846153846154
```

Given the outcome is 0 (not approved) or 1 (approved), we would expect at least a 50% accuracy for this dataset. Thus, the bad quality data is inhibiting the model's ability to detect actual patterns and make informed predictions.

## 4. Reinforcing how bad inputs result in bad outputs

```
In [50]: # Simulate bad input causing unreliable prediction
bad_input = pd.DataFrame({'Age': [999], 'Annual_Income': [1000000], 'Home_Value': [5000000], 'Deposit': [5000], 'Mortgage_Amount': [4995000], 'Credit_Score': [320]})
try:
    print("Prediction:", bad_model.predict(bad_input))
except Exception as e:
    print("Error:", e)
```

```
Prediction: [1]
```

This example demonstrates how a model trained with bad data, predicts that a customer with age 999 and a £500k deposit should be approved for a mortgage.

## Let's try it now with good quality data

### 1. Raw customer data

This data is accurate, and cleaned, with no null values, and stored values aligned with expected values for the attributes.

```
In [51]: # Example raw data from a finance application process
good_data = pd.read_csv(r"C:\Users\GabriellaDixon\Projects\AI Thought Leadership\synthetic_financial_data_good_quality.csv")
good_data
```

|    |    |             |             |             |             |        |   |
|----|----|-------------|-------------|-------------|-------------|--------|---|
| 1  | 42 | 10000.0000  | 518668.1300 | 67246.3500  | 451421.7800 | 322.00 | 0 |
| 2  | 42 | 31929.6400  | 309610.6700 | 63351.3900  | 246259.2800 | 588.00 | 0 |
| 3  | 60 | 68884.2800  | 503270.8300 | 129986.4000 | 373284.4300 | 309.49 | 0 |
| 4  | 55 | 125827.3500 | 205746.3200 | 33830.7600  | 171915.5600 | 568.00 | 0 |
| 5  | 32 | 41804.7500  | 226363.8000 | 17818.5100  | 208545.2900 | 743.00 | 0 |
| 6  | 48 | 136386.2700 | 177935.4100 | 28744.1700  | 149191.2400 | 786.00 | 1 |
| 7  | 57 | 131538.8300 | 406812.6500 | 52182.8800  | 354629.7700 | 737.00 | 0 |
| 8  | 67 | 125874.0200 | 275941.6900 | 75987.5000  | 199954.1900 | 491.00 | 0 |
| 9  | 72 | 0.0000      | 535456.1400 | 126044.2500 | 409411.8900 | 659.00 | 0 |
| 10 | 58 | 45535.2400  | 414954.0500 | 29916.4000  | 385037.6500 | 459.00 | 0 |
| 11 | 80 | 52949.4600  | 227990.5200 | 49366.8600  | 178623.6600 | 744.00 | 1 |
| 12 | 43 | 148677.7904 | 534550.5600 | 87608.1100  | 446942.4500 | 541.00 | 0 |
| 13 | 35 | 43727.5900  | 257598.4200 | 51070.2600  | 206528.1600 | 466.00 | 0 |
| 14 | 32 | 46128.8500  | 370863.6800 | 86716.3400  | 284147.3400 | 458.00 | 0 |
| 15 | 75 | 44851.1300  | 551805.5370 | 104388.9770 | 447416.5600 | 773.00 | 0 |
| 16 | 71 | 39478.4900  | 128519.3025 | 26004.0411  | 102515.2614 | 516.00 | 0 |
| 17 | 42 | 0.0000      | 169759.1700 | 12039.2100  | 157719.9600 | 641.00 | 0 |
| 18 | 50 | 41421.4800  | 358249.5000 | 22276.1400  | 335973.3600 | 464.00 | 0 |
| 19 | 29 | 99519.4600  | 272824.3900 | 40519.9500  | 232304.4400 | 604.00 | 1 |
| 20 | 57 | 52894.5500  | 285494.4100 | 82190.4800  | 203303.9300 | 463.00 | 0 |
| 21 | 18 | 114491.7900 | 498300.9600 | 41574.4864  | 456726.4736 | 539.00 | 0 |
| 22 | 73 | 146567.8200 | 238010.4900 | 40676.7300  | 197333.7600 | 585.00 | 0 |
| 23 | 56 | 77127.5700  | 331923.2300 | 83559.7000  | 248363.5300 | 530.00 | 0 |
| 24 | 63 | 125245.4200 | 539496.3300 | 96450.8300  | 443045.5000 | 805.89 | 1 |
| 25 | 62 | 0.0000      | 179497.6400 | 10400.3300  | 169097.3100 | 676.00 | 0 |
| 26 | 25 | 118860.5900 | 484841.1600 | 64952.5200  | 419888.6400 | 633.00 | 1 |
| 27 | 73 | 60585.4900  | 446107.2900 | 47737.3900  | 398369.9000 | 598.00 | 0 |
| 28 | 38 | 32132.1100  | 430578.3100 | 91301.7400  | 339276.5700 | 630.00 | 0 |
| 29 | 64 | 135877.1600 | 303586.2900 | 20012.8400  | 283573.4500 | 329.00 | 0 |
| 30 | 74 | 96756.8400  | 459123.9100 | 94199.0300  | 364924.8800 | 436.00 | 1 |
| 31 | 80 | 83357.0100  | 328777.5000 | 37634.8000  | 290642.7000 | 659.00 | 0 |
| 32 | 75 | 96636.5800  | 350042.1200 | 29512.3600  | 320529.7600 | 310.00 | 0 |
| 33 | 53 | 14995.5100  | 146658.3900 | 23758.6100  | 122899.7800 | 527.00 | 1 |
| 34 | 50 | 76392.3200  | 140728.3200 | 15025.7800  | 125702.5400 | 656.00 | 1 |
| 35 | 69 | 31680.8200  | 305074.0100 | 23341.5100  | 281732.5000 | 786.00 | 0 |
| 36 | 28 | 136658.1400 | 172165.4700 | 10499.7000  | 155665.7700 | 545.00 | 0 |
| 37 | 45 | 121028.7400 | 154795.6800 | 10899.1200  | 144706.7600 | 434.00 | 0 |
| 38 | 66 | 46792.5800  | 246365.0400 | 18031.3200  | 225653.7200 | 534.00 | 1 |
| 39 | 42 | 118225.1400 | 151742.7700 | 37499.6200  | 114243.1500 | 638.00 | 1 |
| 40 | 58 | 78112.0000  | 145731.3700 | 9624.8700   | 136106.5000 | 509.00 | 0 |
| 41 | 18 | 0.0000      | 448166.8020 | 97413.7400  | 340733.1800 | 381.00 | 0 |
| 42 | 55 | 36403.1200  | 519518.5700 | 75858.7200  | 443659.8500 | 538.00 | 0 |
| 43 | 37 | 114648.6500 | 346068.5200 | 80895.9000  | 265183.4300 | 724.00 | 1 |
| 44 | 42 | 124975.3600 | 171239.1200 | 2125.7600   | 149073.3600 | 326.00 | 0 |
| 45 | 64 | 148447.2500 | 253142.7500 | 66268.6000  | 196874.1500 | 505.00 | 0 |
| 46 | 18 | 110348.9000 | 482116.8800 | 31918.8400  | 450198.0500 | 330.00 | 0 |
| 47 | 54 | 122633.8900 | 181238.2000 | 42329.3500  | 138908.9500 | 700.00 | 1 |
| 48 | 36 | 131402.3200 | 433885.5900 | 79947.2900  | 353938.3000 | 473.00 | 0 |
| 49 | 54 | 0.0000      | 509172.6900 | 88486.4200  | 420686.2700 | 334.00 | 0 |
| 50 | 49 | 109528.0100 | 479489.4500 | 128519.5200 | 350699.9300 | 748.00 | 1 |

