



```
[2]: df = pd.read_csv('/Users/rohitkumar/Downloads/spam.csv', encoding='latin1')
```

```
[3]: df.sample(5)
```

| [3]: | v1        | v2  | Unnamed: 2 | Unnamed: 3 | Unnamed: 4 |
|------|-----------|---|------------|------------|------------|
|      | 1676 ham  | Yeah, that's fine! It's å£6 to get in, is that... | NaN        | NaN        | NaN        |
|      | 4926 ham  | Wanna do some art?! :D                            | NaN        | NaN        | NaN        |
|      | 3370 ham  | Tomorrow i am not going to theatre. . . So i c... | NaN        | NaN        | NaN        |
|      | 2945 ham  | make that 3! 4 fucks sake?! x                     | NaN        | NaN        | NaN        |
|      | 3389 spam | Please CALL 08712402972 immediately as there i... | NaN        | NaN        | NaN        |

```
[4]: df.shape
```

```
[4]: (5572, 5)
```

```
[5]: # 1. Data cleaning
# 2. EDA
# 3. Text Preprocessing
# 4. Model building
# 5. Evaluation
# 6. Improvement
# 7. Website
# 8. Deploy
```

## 1. Data Cleaning

```
[6]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5572 entries, 0 to 5571
Data columns (total 5 columns):
#   Column      Non-Null Count  Dtype
---  -
0    v1          5572 non-null  object
1    v2          5572 non-null  object
2    Unnamed: 2  50 non-null    object
```