Data Collection

```
import numpy as np
import pandas as pd
import matplotlib as mpl
import matplotlib.pyplot as plt
%matplotlib inline
%config InlineBackend.figure_fmt = 'svg'
```

```
df_ipo.head.iloc['Daily Offers.csv']
```

	ipo_date	company name	customer	country	status	Days application	thickness	width	material_id	product_id	Delivery date	selling price
0	01/04/2016 00:00:00	ECOPRINT ROTO-GRAVURE PRIVATE LIMITED	2011011130	01/04/2016	01/04/2016	0	0.5	1.25	2.00	2.00	15/01/1900 00:00:00	2017/07/31	2016-04-01
1	01/04/2016 00:00:00	ECOPRINT ROTO-GRAVURE PRIVATE LIMITED	2011011130	01/04/2016	01/04/2016	0	0.5	1.25	2.00	2.00	15/01/1900 00:00:00	2017/07/31	2016-04-01
2	01/04/2016 00:00:00	SUMITOMO CORPORATION INDIA PVT. LTD.	2011011130	01/04/2016	01/04/2016	0	0.5	1.25	2.00	2.00	15/01/1900 00:00:00	2017/07/31	2016-04-01
3	01/04/2016 00:00:00	TRIDENT PACKAGING INDIA PVT. LTD.	2011011130	01/04/2016	01/04/2016	0	0.5	1.25	2.00	2.00	15/01/1900 00:00:00	2017/07/31	2016-04-01
4	01/04/2016 00:00:00	ALLIED PACKAGING INDUSTRIES PVT. LTD.	2011011130	01/04/2016	01/04/2016	0	0.5	1.25	2.00	2.00	15/01/1900 00:00:00	2017/07/31	2016-04-01

```
df_ipo.shape
```

```
(2675, 12)
```

	ipo_date	company name	customer	country	status	Days application	thickness	width	material_id	product_id	Delivery date	selling price
2670	01/04/2016 00:00:00	ECOPRINT ROTO-GRAVURE PRIVATE LIMITED	2011011130	01/04/2016	01/04/2016	0	0.5	1.25	2.00	2.00	15/01/1900 00:00:00	2017/07/31	2016-04-01
2671	01/04/2016 00:00:00	ECOPRINT ROTO-GRAVURE PRIVATE LIMITED	2011011130	01/04/2016	01/04/2016	0	0.5	1.25	2.00	2.00	15/01/1900 00:00:00	2017/07/31	2016-04-01
2672	01/04/2016 00:00:00	ECOPRINT ROTO-GRAVURE PRIVATE LIMITED	2011011130	01/04/2016	01/04/2016	0	0.5	1.25	2.00	2.00	15/01/1900 00:00:00	2017/07/31	2016-04-01
2673	01/04/2016 00:00:00	ECOPRINT ROTO-GRAVURE PRIVATE LIMITED	2011011130	01/04/2016	01/04/2016	0	0.5	1.25	2.00	2.00	15/01/1900 00:00:00	2017/07/31	2016-04-01
2674	01/04/2016 00:00:00	ALLIED PACKAGING INDUSTRIES PVT. LTD.	2011011130	01/04/2016	01/04/2016	0	0.5	1.25	2.00	2.00	15/01/1900 00:00:00	2017/07/31	2016-04-01

```
DataFrame_ipo.columns
```

```
Index(['ipo_date', 'company name', 'customer', 'country', 'status',
       'Days application', 'thickness', 'width', 'material_id',
       'product_id', 'Delivery date', 'selling price'],
      dtype='object')
```

```
DataFrame_ipo.shape
```

```
(2675, 12)
```

Exploring Data Analysis

```
df_info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2675 entries, 0 to 2674
Data columns (total 12 columns):
#   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   ipo_date          2675 non-null   object
 1   company name      2675 non-null   object
 2   customer          2675 non-null   int64
 3   country           2675 non-null   int64
 4   status            2675 non-null   object
 5   Days application  2675 non-null   int64
 6   thickness         2675 non-null   float64
 7   width             2675 non-null   int64
 8   material_id       2675 non-null   int64
 9   product_id        2675 non-null   int64
 10  Delivery date     2675 non-null   int64
 11  selling price     2675 non-null   float64
dtypes: float64(2), int64(7), object(3)
memory usage: 247.8 KB
```

```
df_info().head()
```

```
ipo_date                 object
company name             object
customer                  int64
country                   int64
status                   object
Days application          int64
thickness                float64
width                     int64
material_id               int64
product_id                int64
Delivery date             int64
selling price            float64
dtypes: object(3), int64(7), float64(2)
memory usage: 247.8 KB
```

```
df_info().tail()
```

```
ipo_date                 object
company name             object
customer                  int64
country                   int64
status                   object
Days application          int64
thickness                float64
width                     int64
material_id               int64
product_id                int64
Delivery date             int64
selling price            float64
dtypes: object(3), int64(7), float64(2)
memory usage: 247.8 KB
```

```
df_describe()
```

	customer	country	application	thickness	width	product_id	delivery date	selling price
count	2.675000e+03	2675.000000	2675.000000	2675.000000	2675.000000	2675.000000	2.675000e+03	2.675000e+03
mean	2.011011e+09	1.000000	0.500000	1.250000	1.250000	2.000000	2.017080e+07	2.016504e+03
std	1.551124e+02	0.000000	0.000000	0.000000	0.000000	0.000000	1.551124e+02	1.551124e+02
min	2.011011e+09	1.000000	0.500000	1.250000	1.250000	2.000000	2.017080e+07	2.016504e+03
25%	2.011011e+09	1.000000	0.500000	1.250000	1.250000	2.000000	2.017080e+07	2.016504e+03
50%	2.011011e+09	1.000000	0.500000	1.250000	1.250000	2.000000	2.017080e+07	2.016504e+03
75%	2.011011e+09	1.000000	0.500000	1.250000	1.250000	2.000000	2.017080e+07	2.016504e+03
max	2.011011e+09	1.000000	0.500000	1.250000	1.250000	2.000000	2.017080e+07	2.016504e+03

Removing the Null Values

```
df_isnull()["customer"].sum()
```

```
0
```

```
df_isnull().T[customer].sum(axis=0)
```

```
0
```

```
df_isnull().head()
```

```
ipo_date            0
company name        0
customer            0
country             0
status              0
Days application    0
thickness           0
width               0
material_id         0
product_id          0
Delivery date       0
selling price       0
dtype: int64
```

```
df_isnull().tail()
```

```
ipo_date            0
company name        0
customer            0
country             0
status              0
Days application    0
thickness           0
width               0
material_id         0
product_id          0
Delivery date       0
selling price       0
dtype: int64
```

```
df_isnull()["customer"].sum(axis=0)
```

```
0
```

```
df_isnull().head()
```

```
ipo_date            0
company name        0
customer            0
country             0
status              0
Days application    0
thickness           0
width               0
material_id         0
product_id          0
Delivery date       0
selling price       0
dtype: int64
```

```
df_isnull()["customer"].sum(axis=0)
```

```
0
```

```
df_isnull().head()
```

```
ipo_date            0
company name        0
customer            0
country             0
status              0
Days application    0
thickness           0
width               0
material_id         0
product_id          0
Delivery date       0
selling price       0
dtype: int64
```

```
df_isnull().tail()
```

```
ipo_date            0
company name        0
customer            0
country             0
status              0
Days application    0
thickness           0
width               0
material_id         0
product_id          0
Delivery date       0
selling price       0
dtype: int64
```

```
df_isnull()["selling price"].sum(axis=0)
```

```
0
```

```
df_isnull().T[selling price].sum(axis=0)
```

```
0
```

```
df_info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2675 entries, 0 to 2674
Data columns (total 12 columns):
#   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   ipo_date          2675 non-null   object
 1   company name      2675 non-null   object
 2   customer          2675 non-null   int64
 3   country           2675 non-null   int64
 4   status            2675 non-null   object
 5   Days application  2675 non-null   int64
 6   thickness         2675 non-null   float64
 7   width             2675 non-null   int64
 8   material_id       2675 non-null   int64
 9   product_id        2675 non-null   int64
 10  Delivery date     2675 non-null   int64
 11  selling price     2675 non-null   float64
dtypes: float64(2), int64(7), object(3)
memory usage: 247.8 KB
```

```
df_info().T[selling price].sum(axis=0)
```

```
nan
```

```
df_info().T[selling price].sum(axis=0).T
```

```
nan
```

```
df_info().T[selling price].sum(axis=0).T.replace(nan,0)
```

```
0.0
```

```
df_info().T[selling price].sum(axis=0).T.replace(nan,0).T
```

```
0.0
```

```
df_info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2675 entries, 0 to 2674
Data columns (total 12 columns):
#   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   ipo_date          2675 non-null   object
 1   company name      2675 non-null   object
 2   customer          2675 non-null   int64
 3   country           2675 non-null   int64
 4   status            2675 non-null   object
 5   Days application  2675 non-null   int64
 6   thickness         2675 non-null   float64
 7   width             2675 non-null   int64
 8   material_id       2675 non-null   int64
 9   product_id        2675 non-null   int64
 10  Delivery date     2675 non-null   int64
 11  selling price     2675 non-null   float64
dtypes: float64(2), int64(7), object(3)
memory usage: 247.8 KB
```

```
df_info().T[selling price].sum(axis=0).T.replace(nan,0).T
```

```
0.0
```

```
df_info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2675 entries, 0 to 2674
Data columns (total 12 columns):
#   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   ipo_date          2675 non-null   object
 1   company name      2675 non-null   object
 2   customer          2675 non-null   int64
 3   country           2675 non-null   int64
 4   status            2675 non-null   object
 5   Days application  2675 non-null   int64
 6   thickness         2675 non-null   float64
 7   width             2675 non-null   int64
 8   material_id       2675 non-null   int64
 9   product_id        2675 non-null   int64
 10  Delivery date     2675 non-null   int64
 11  selling price     2675 non-null   float64
dtypes: float64(2), int64(7), object(3)
memory usage: 247.8 KB
```

```
df_info()["status"].replace(0,"Nan")
```

```
NaN    Nan
Name: status, dtype: object
```

```
df_info()["status"].replace(0,"Nan").T
```

```
NaN    Nan
Name: status, dtype: object
```

```
df_info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2675 entries, 0 to 2674
Data columns (total 12 columns):
#   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   ipo_date          2675 non-null   object
 1   company name      2675 non-null   object
 2   customer          2675 non-null   int64
 3   country           2675 non-null   int64
 4   status            2675 non-null   object
 5   Days application  2675 non-null   int64
 6   thickness         2675 non-null   float64
 7   width             2675 non-null   int64
 8   material_id       2675 non-null   int64
 9   product_id        2675 non-null   int64
 10  Delivery date     2675 non-null   int64
 11  selling price     2675 non-null   float64
dtypes: float64(2), int64(7), object(3)
memory usage: 247.8 KB
```

```
df_info()["status"].replace(0,"Nan").T
```

```
NaN    Nan
Name: status, dtype: object
```

Splitting The Data

```
df_info[['ipo_', 'ipo']]
df_info[['ipo']]
```

```
.
```

	ipo_date	customer	country	application	thickness	width	product_id	delivery date	selling price
0	2016/04/01	2011011130	1	0.5	1.25	2.00	2.00	2017/07/31	2016-04-01
1	2016/04/01	2011011130	1	0.5	1.25	2.00	2.00	2017/07/31	2016-04-01
2	2016/04/01	2011011130	1	0.5	1.25	2.00	2.00	2017/07/31	2016-04-01
3	2016/04/01	2011011130	1	0.5	1.25	2.00	2.00	2017/07/31	2016-04-01
4	2016/04/01	2011011130	1	0.5	1.25	2.00	2.00	2017/07/31	2016-04-01
2670	2016/04/01	2011011130	1	0.5	1.25	2.00	2.00	2017/07/31	2016-04-01
2671	2016/04/01	2011011130	1	0.5	1.25	2.00	2.00	2017/07/31	2016-04-01
2672	2016/04/01	2011011130	1	0.5	1.25	2.00	2.00	2017/07/31	2016-04-01
2673	2016/04/01	2011011130	1	0.5	1.25	2.00	2.00	2017/07/31	2016-04-01
2674	2016/04/01	2011011130	1	0.5	1.25	2.00	2.00	2017/07/31	2016-04-01

```
df_info()[['ipo_', 'ipo']]
```

```
2675 rows × 12 columns
```

```
y
```

```
ipo_date    object
customer    int64
country     int64
status      object
Days application    int64
thickness   float64
width       int64
material_id         int64
product_id          int64
Delivery date       int64
selling price       float64
dtype: object, int64
```

```
df_info()[['ipo', 'customer']]
```

```
2675 rows × 2 columns
```

```
y
```

```
ipo_date    object
customer    int64
dtype: object, int64
```

```
df_info()[['ipo']]
```

```
2675 rows × 1 columns
```

```
y
```

```
ipo_date    object
dtype: object
```

Data Visualization

```
df_info()[['ipo']].hist()
```

```
<matrix_ref_1>
```