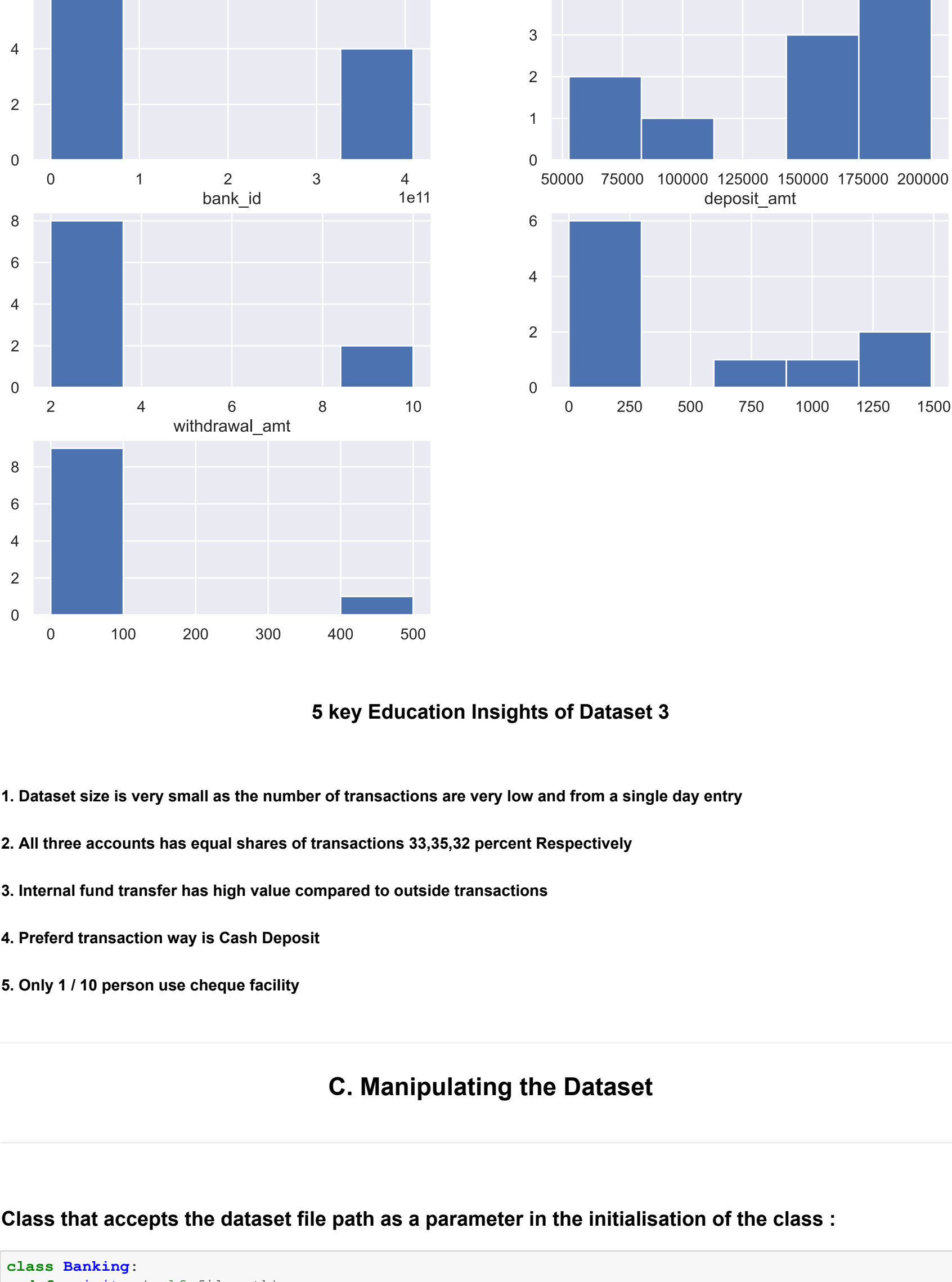






```
dataset3.df.hist(bins=5)
plt.show()
```



### 5 key Education Insights of Dataset 3

1. Dataset size is very small as the number of transactions are very low and from a single day entry
2. All three accounts has equal shares of transactions 33,35,32 percent Respectively
3. Internal fund transfer has high value compared to outside transactions
4. Preferd transaction way is Cash Deposit
5. Only 1 / 10 person use cheque facility

## C. Manipulating the Dataset

**Class that accepts the dataset file path as a parameter in the initialisation of the class :**

```
In [189]: class Banking:
def __init__(self,filepath):
    try:
        self.path_Dataset = filepath
        self.df = pd.read_excel(self.path_Dataset)
        # self.df = self.df.dropna()
        print("\nColumn Names found : \n",self.df.columns)
    except:
        print("An exception occurred try again with different parameters")

def stdDev(self,string):
    try:
        # print(self.df[string])
        print("\nStandard Deviation : ")
        print(np.std(self.df[string].fillna(0)))
    except:
        print("An exception occurred ")

def mean(self,string):
    try:
        # print("\nMean : ")
        print(np.mean(self.df[string].fillna(0)))
    except:
        print("An exception occurred ")

def outliers(self,string):
    try:
        # print("\nOutliers :\n")
        z_scores = stats.zscore(self.df[string].fillna(0))
        abs_z_scores = np.abs(z_scores)
        filtered_entries = (abs_z_scores > 3)
        print(self.df[filtered_entries])
    except:
        print("An exception occurred ")

print("Dataset paths:\n")
print("\n1. https://drive.google.com/uc?export=download&id=1QCk_te229t56RkFt2RkzU58FD2kedg1Z")
print("\n2. https://drive.google.com/uc?export=download&id=1fkTQW4w1ih9j91c52CNcDv8XNdJ3KMJ_")

Dataset paths:

1. https://drive.google.com/uc?export=download&id=1QCk_te229t56RkFt2RkzU58FD2kedg1Z

2. https://drive.google.com/uc?export=download&id=1fkTQW4w1ih9j91c52CNcDv8XNdJ3KMJ_

=== Initialization with dataset path as parameter ===

(Use path string from above list)
```

```
In [189]: obj = Banking('https://drive.google.com/uc?export=download&id=1QCk_te229t56RkFt2RkzU58FD2kedg1Z')

Column Names found :
Index(['Account No.', 'DATE', 'TRANSACTION DETAILS', 'CHQ.NO.', 'VALUE DATE',
      'WITHDRAWAL AMT', 'DEPOSIT AMT', 'BALANCE AMT', '.'],
      dtype='object')

=== below are the functions as per requested with peticular parameter ===

In [190]: # Insert Column name of your choice from above list as afunction parameter
#function for standard deviation
obj.stdDev("WITHDRAWAL AMT")

#function for Mean
obj.mean("WITHDRAWAL AMT")

#function for Outliers / Z-score
obj.outliers("WITHDRAWAL AMT")
# Parameters added to reduce the execution time as our 2nd dataset is large and taking too much time to
calculate z-score

Standard Deviation :
7696848.178413423

Mean :
2068757.001099466

Outliers :

Account No      DATE      TRANSACTION DETAILS  CHQ.NO.  \
2926  409000425051'  2018-10-31      TRF TO Myur Joshi      NaN
2939  409000405747'  2016-01-30      TRF TO Haur Joshi      NaN
2940  409000405747'  2016-01-30      TRF TO Indiaforensic SERVICES IN  NaN
2972  409000405747'  2018-09-28      Payments For : 9090000577  NaN
2991  409000438611'  2016-02-26      TRF TO M Joshi          NaN
...
109190 409000362497'  2017-11-20      IMPS 191117 1C          NaN
109368 409000362497'  2017-11-27      IMPS 251117 1C          NaN
109707 409000362497'  2017-12-11      IMPS 091217 1C          NaN
109755 409000362497'  2017-12-12      IMPS 121217 3C          NaN
109863 409000362497'  2017-12-18      IMPS 171217 1C          NaN

VALUE DATE  WITHDRAWAL AMT  DEPOSIT AMT  BALANCE AMT  .
2926  2018-10-31      3.540000e+08      NaN -3.535408e+08  .
2939  2016-01-30      1.700000e+08      NaN -5.267348e+08  .
2940  2016-01-30      3.000000e+07      NaN -5.567348e+08  .
2972  2018-09-28      1.030203e+08      NaN -5.467348e+08  .
2991  2016-02-26      1.300000e+08      NaN -5.080175e+08  .
...
109190 2017-11-20      2.798619e+07      NaN -1.901105e+09  .
109368 2017-11-27      4.008094e+07      NaN -1.898634e+09  .
109707 2017-12-11      3.640580e+07      NaN -1.897443e+09  .
109755 2017-12-12      2.460039e+07      NaN -1.895761e+09  .
109863 2017-12-18      3.341938e+07      NaN -1.939116e+09  .

[1766 rows x 9 columns]
```

### Questions:

#### 1. What (if any) experience do you have in Cloud Services (eg. AWS, GCP)?

Ans : I have undergone through online certification for AWS on DataCamp/Udemy platform. I do not hold any professional experience on AWS or GCP But I have basic knowledge of AWS concepts and deploying applications on cloud. I would love to learn more about it professionally in near future. Certification Link : <https://www.datacamp.com/certificate-of-accomplishment/course/24f90a9a956ce20623f145457bac212a0db1dd5>

#### 2. What is your experience with building machine learning models? What packages are you familiar with?

Ans : By being a student of MSc Data Science and 2+ years of diverse-industry experience, I am able to analyze the input dataset and implement the machine learning models thoroughly. I also possesses knowledge of machine learning in R and Python both

Some of the packages as below: Matplotlib, Pandas, NumPy, SciPy, NLTK, Scikit-learn, Statsmodels, Seaborn, ggplot2, mir, superml, etc. I am aware of classification and clustering algorithms and implementation for the same. Also undergoing the practicals on Keras and TensorFlow libraries.

Besides it, I am actively doing freelancing on machine learning projects. Which consists delivering projects starting from analyzing the dataset, data preprocessing, EDA, Statistical Models along with the Data Visualization techniques.

**Thank You!**

**Kindly revert back in case of queries.**