Business Context:

A Bank wants to develop a customer segmentation to define marketing strategy. The sample dataset summarizes the usage behaviour of about 9000 active credit card holders during the last 6 months. The file at customer level includes Transaction frequency, amount, tenure etc.

Business Requirements:

Bank Marketing team would like to leverage ML to launch target ad campaign that is tailored to specific group of customers.Based on this group or Market segments, marketing strategy will be formed.

Following Parameters would help maketing team to formulate better strategy,

- Monthly average spend
- Purchases type (EMI etc)
- Average amount per purchase.
- Clustering segments of credit card holders
- Provide the strategic insights and implementation of strategies for given set of cluster characteristics.

Data Dictionary:

- CUST_ID: Credit card holder ID
- BALANCE: Monthly average balance (based on daily balance averages)
- BALANCE_FREQUENCY: Ratio of last 12 months with balance
- PURCHASES: Total purchase amount spent during last 12 months
- ONEOFF_PURCHASES: Total amount of one-off purchases
- INSTALLMENTS_PURCHASES: Total amount of installment purchases
- CASH_ADVANCE: Total cash-advance amount
- PURCHASES_ FREQUENCY: Frequency of purchases (Percent of months with at least one purchase)
- ONEOFF_PURCHASES_FREQUENCY: Frequency of one-off-purchases
PURCHASES_INSTALLMENTS_FREQUENCY: Frequency of installment purchases
- CASH*ADVANCE* FREQUENCY: Cash-Advance frequency
- AVERAGE_PURCHASE_TRX: Average amount per purchase transaction
- CASH_ADVANCE_TRX: Average amount per cash-advance transaction
- PURCHASES_TRX: Average amount per purchase transaction
- CREDIT_LIMIT: Credit limit
- PAYMENTS: Total payments (due amount paid by the customer to decrease their statement balance) in the period
- MINIMUM_PAYMENTS: Total minimum payments due in the period.
- PRC_FULL_PAYMEN: Percentage of months with full payment of the due statement balance
- TENURE: Number of months as a customer

Methodology

- Exploratory Analysis to describe the data
- K Means Clustering to Segment Customer

Exploratory Analysis

In [65]:
```
print('Average Monthly Purchase Amount is: ',credit['Monthly_avg_purchase'].mean(),'\n '
,
      'Minimum Monthly Purchase Amount is: ',credit['Monthly_avg_purchase'].min(),'\n',
      'Maximum Monthly Purchase Amount is: ',credit['Monthly_avg_purchase'].max(),'\n',
      'Average Customer Tenure (in Months) is: ', credit['TENURE'].mean(),'\n',
      'Total Amount of Installment Purchases is: ', credit['INSTALLMENTS_PURCHASES'].sum
(),'\n',
      'Total Amount of One Off Purchases is : ', credit['ONEOFF_PURCHASES'].sum())
```

```
Average Monthly Purchase Amount is:  86.1751728841086
  Minimum Monthly Purchase Amount is:  0.0
 Maximum Monthly Purchase Amount is:  4086.630833333333
 Average Customer Tenure (in Months) is:  11.51731843575419
 Total Amount of Installment Purchases is:  3679055.42
 Total Amount of One Off Purchases is :  5302314.470000001
```

- Amount of purchases on Installment are much lower than One Off Purchases
- There are customer who are not using CC

Customer Spending Behaviour

In [44]:
```
credit['purchase_type'].value_counts()
```

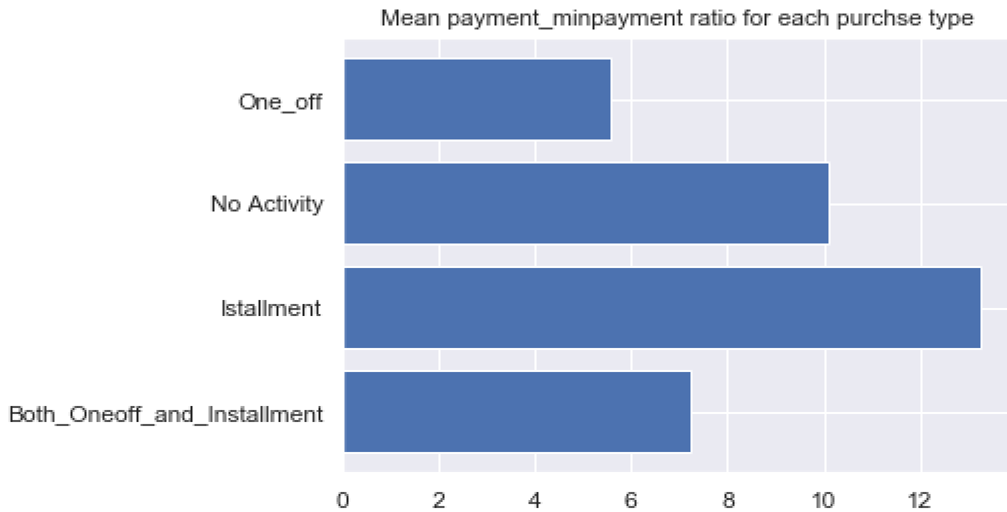
Out[44]:
```
Both_Oneoff_and_Installment    2774
Istallment                     2260
No Activity                    2042
One_off                        1874
Name: purchase_type, dtype: int64
```

- There are large set of Customer that are not using Credit Card for Purchase (may be using only for cash advance).
- There are Customers that Perform Both Oneoff and Installment
- Approx. 1800 Subs do not prefer Installment based transactions.

Average Payment to Minimum payment ratio for each purchase type.

In [51]:
```
fig,ax=plt.subplots()
ax.barh(y=range(len(x)), width=x.values,align='center')
ax.set(yticks= np.arange(len(x)),yticklabels = x.index);
plt.title('Mean payment_minpayment ratio for each purchse type')
```

Out[51]:
```
Text(0.5, 1.0, 'Mean payment_minpayment ratio for each purchse type')
```



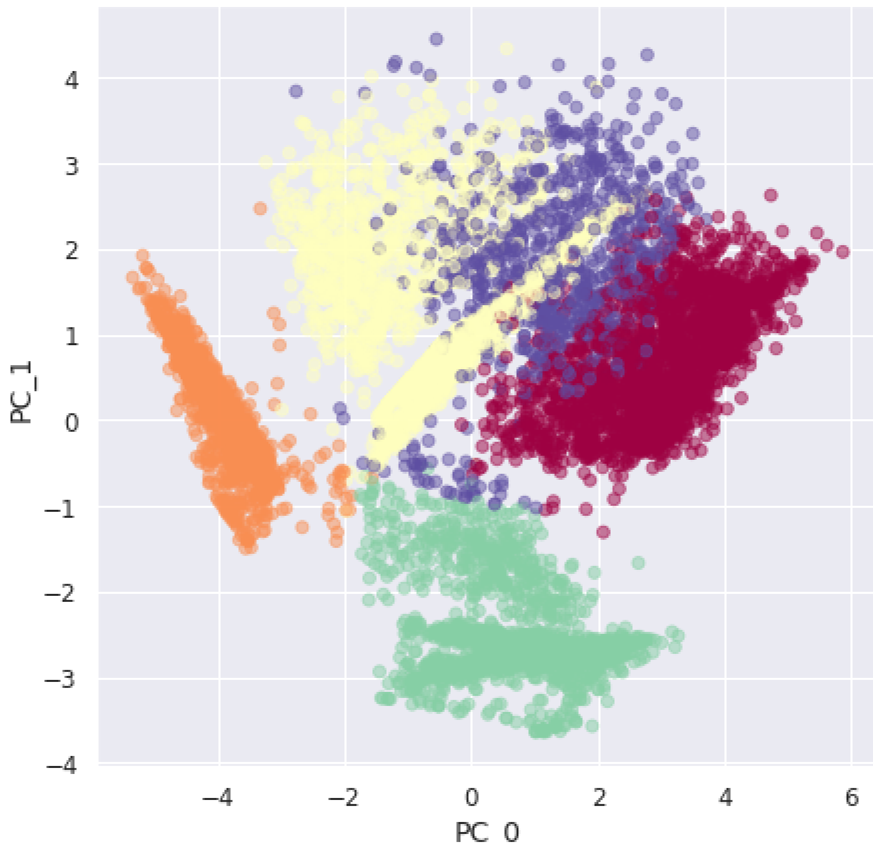
Clustering Analysis with K Means

Clustering

- Clustering with K as 5

In [50]:
```
plt.figure(figsize=(7,7))
plt.scatter(reduced_cr[:,0],reduced_cr[:,1],c=km_5.labels_,cmap='Spectral',alpha=0.5)
plt.xlabel('PC_0')
plt.ylabel('PC_1')
```

Out[50]: Text(0, 0.5, 'PC_1')



In [58]:
```
# Finding Mean of features for each cluster
cluster_5 = cluster_df_5.groupby('Cluster_5')\
.apply(lambda x: x[col_kpi].mean()).T
cluster_5
```

Out[58]:

Cluster_5	0	1	2	3	4
PURCHASES_TRX	34.501008	0.024120	7.086910	11.917011	28.245283
Monthly_avg_purchase	209.169344	0.064284	68.846861	47.287449	147.026040
Monthly_cash_advance	3.760209	185.211776	75.993788	23.616800	249.750029
CASH_ADVANCE_TRX	0.142137	6.446213	2.773605	0.665750	10.389151
payment_minpay	8.500536	9.990736	5.570556	13.611466	4.048983
both_oneoff_installment	1.000000	0.000000	0.000536	0.000000	0.930425
istallment	0.000000	0.013025	0.000000	1.000000	0.061321
one_off	0.000000	0.001930	0.999464	0.000000	0.008255
none	0.000000	0.985046	0.000000	0.000000	0.000000
CREDIT_LIMIT	5679.139581	4049.622126	4501.521004	3251.863083	5988.679245

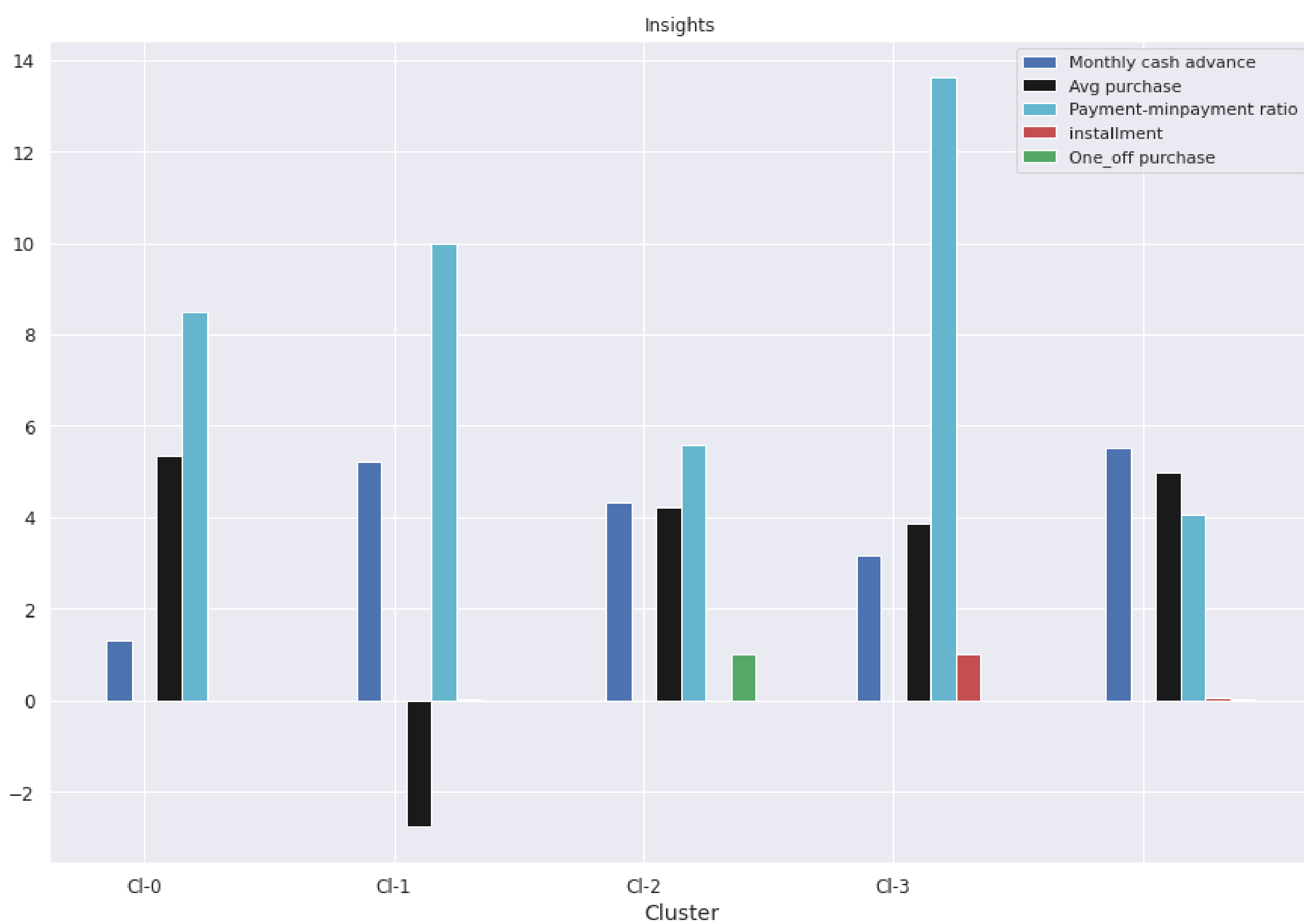
```
In [61]: fig,ax=plt.subplots(figsize=(15,10))
         index=np.arange(len(cluster_5.columns))

         cash_advance=np.log(cluster_5.loc['Monthly_cash_advance',:].values)
         #credit_score=(cluster_5.loc['limit_usage',:].values)
         purchase= np.log(cluster_5.loc['Monthly_avg_purchase',:].values)
         payment=cluster_5.loc['payment_minpay',:].values
         installment=cluster_5.loc['istallment',:].values
         one_off=cluster_5.loc['one_off',:].values


         bar_width=.10
         b1=plt.bar(index,cash_advance,color='b',label='Monthly cash advance',width=bar_width)
         #b2=plt.bar(index+bar_width,credit_score,color='m',label='Credit_score',width=bar_width)
         b3=plt.bar(index+2*bar_width,purchase,color='k',label='Avg purchase',width=bar_width)
         b4=plt.bar(index+3*bar_width,payment,color='c',label='Payment-minpayment ratio',width=ba
         r_width)
         b5=plt.bar(index+4*bar_width,installment,color='r',label='installment',width=bar_width)
         b6=plt.bar(index+5*bar_width,one_off,color='g',label='One_off purchase',width=bar_width)

         plt.xlabel("Cluster")
         plt.title("Insights")
         plt.xticks(index + bar_width, ('Cl-0', 'Cl-1', 'Cl-2', 'Cl-3'))
         plt.legend()
```

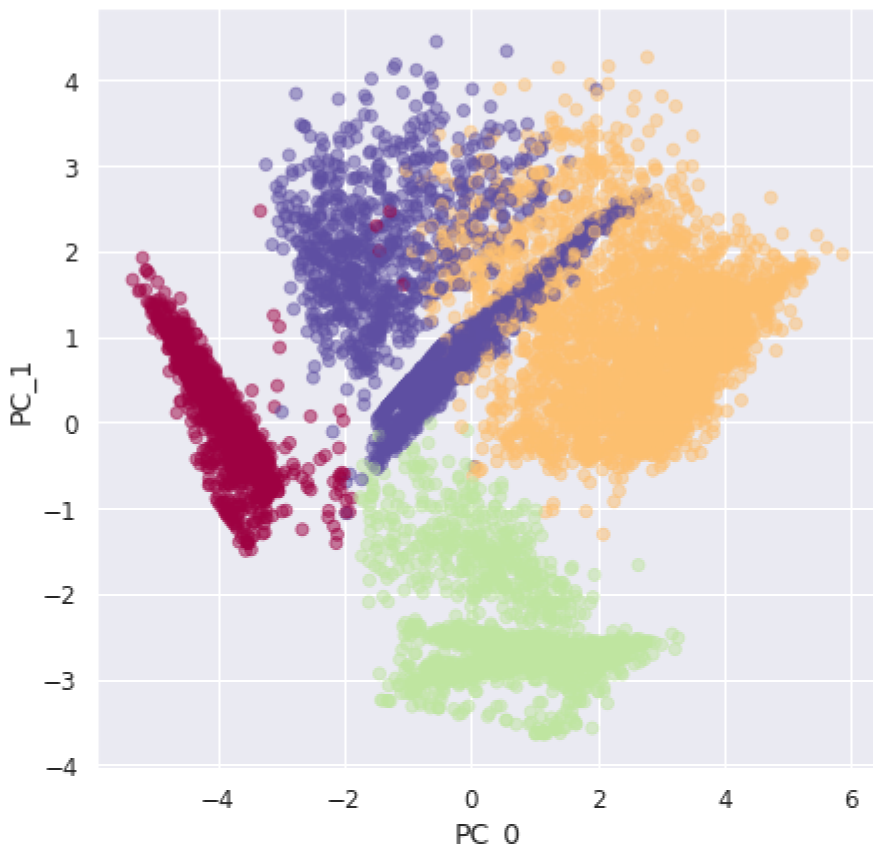
```
Out[61]: <matplotlib.legend.Legend at 0x7f6e893ee400>
```



Performing for k=4

In [63]:
```
plt.figure(figsize=(7,7))
plt.scatter(reduced_cr[:,0],reduced_cr[:,1],c=km_4.labels_,cmap='Spectral',alpha=0.5)
plt.xlabel('PC_0')
plt.ylabel('PC_1')
```

Out[63]:
```
Text(0, 0.5, 'PC_1')
```



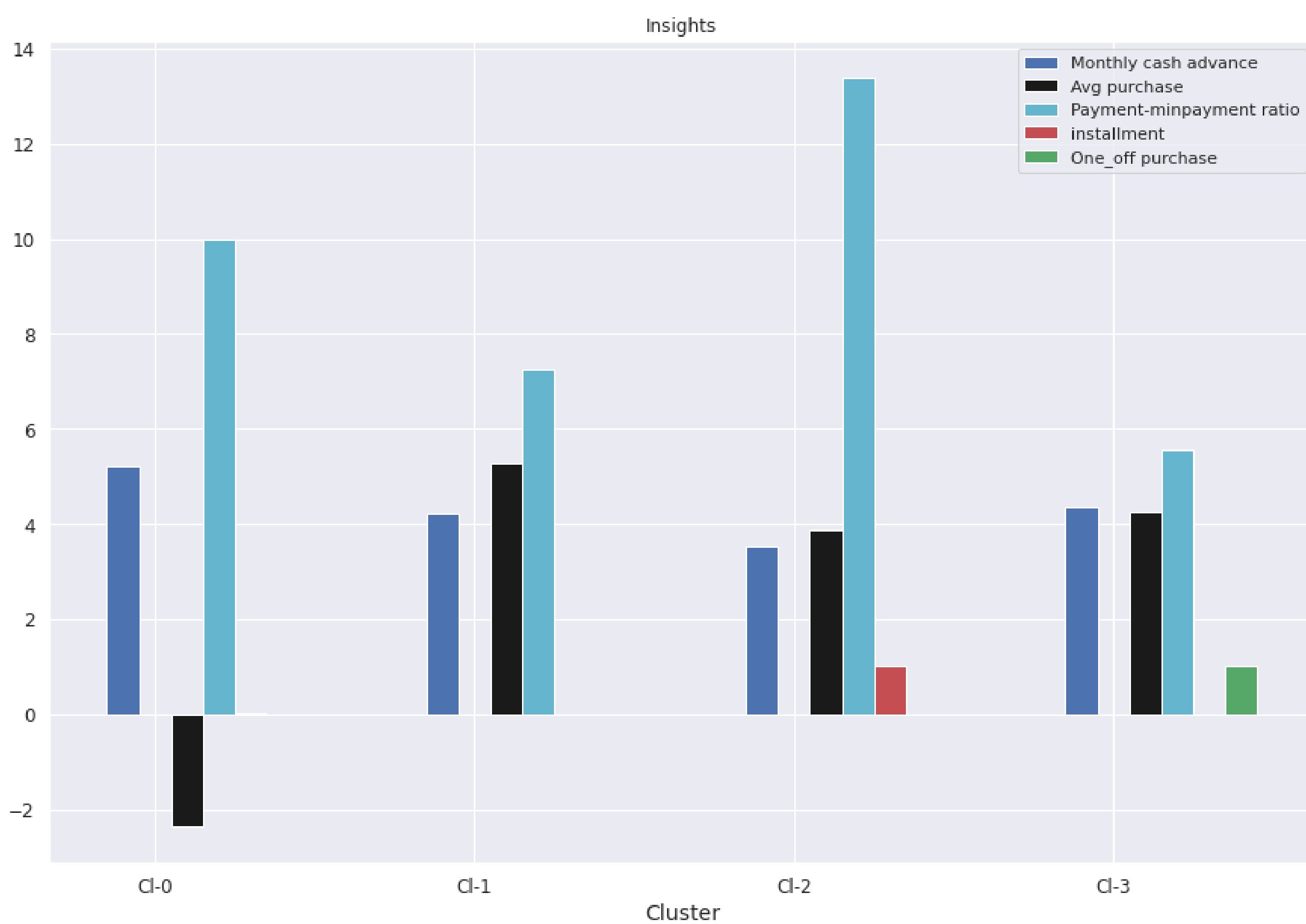
In [66]:
```
fig,ax=plt.subplots(figsize=(15,10))
index=np.arange(len(cluster_4.columns))

cash_advance=np.log(cluster_4.loc['Monthly_cash_advance',:].values)
#credit_score=(cluster_4.loc['limit_usage',:].values)
purchase= np.log(cluster_4.loc['Monthly_avg_purchase',:].values)
payment=cluster_4.loc['payment_minpay',:].values
installment=cluster_4.loc['istallment',:].values
one_off=cluster_4.loc['one_off',:].values


bar_width=.10
b1=plt.bar(index,cash_advance,color='b',label='Monthly cash advance',width=bar_width)
#b2=plt.bar(index+bar_width,credit_score,color='m',label='Credit_score',width=bar_width)
b3=plt.bar(index+2*bar_width,purchase,color='k',label='Avg purchase',width=bar_width)
b4=plt.bar(index+3*bar_width,payment,color='c',label='Payment-minpayment ratio',width=ba
r_width)
b5=plt.bar(index+4*bar_width,installment,color='r',label='installment',width=bar_width)
b6=plt.bar(index+5*bar_width,one_off,color='g',label='One_off purchase',width=bar_width)

plt.xlabel("Cluster")
plt.title("Insights")
plt.xticks(index + bar_width, ('Cl-0', 'Cl-1', 'Cl-2', 'Cl-3'))
plt.legend()
```

Out[66]: <matplotlib.legend.Legend at 0x7f6e892d9c88>



Insights with 4 Clusters

- Cluster 2 is the group of customers who have highest Monthly_avg purchases and doing both installment as well as one_off purchases, have comparatively good credit score. This group is about 31% of the total customer base
- cluster 1 is taking maximum advance_cash and is paying comparatively less minimum payment and poor credit_score & doing no purchase transaction. This group is about 23% of the total customer base
- Cluster 0 customers are doing maximum One_Off transactions and least payment ratio and credit_score on lower side This group is about 21% of the total customer base
- Cluster 3 customers have maximum credit score and are paying dues and are doing maximum installment purchases. This group is about 25% of the total customer base

Conclusion

Marketing Strategy Suggested:

- a. Group 2 They are potential target customers who are paying dues and doing purchases and maintaining comparatively good credit score) -- we can increase credit limit or can lower down interest rate -- Can be given premium card /loyality cards to increase transactions
- b. Group 1 They have poor credit score and taking only cash on advance. We can target them by providing less interest rate on purchase transaction
- c. Group 0 This group is has minimum paying ratio and using card for just oneoff transactions (may be for utility bills only). This group seems to be risky group.
- d. Group 3 This group is performing best among all as cutomers are maintaining good credit score and paying dues on time. -- Giving rewards point will make them perform more purchases.