

In [103]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

In [104]:

```
card = pd.read_csv('training.csv')
```

In [105]:

```
card.head()
```

Out[105]:

	Unnamed: 0	SeriousDlqin2yrs	RevolvingUtilizationOfUnsecuredLines	age	NumberOfTimes59DaysPastDueNo
0	1	1	0.766127	45	
1	2	0	0.957151	40	
2	3	0	0.658180	38	
3	4	0	0.233810	30	
4	5	0	0.907239	49	

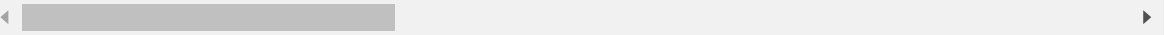
In [106]:

```
card.isnull()
```

Out[106]:

	Unnamed: 0	SeriousDlqin2yrs	RevolvingUtilizationOfUnsecuredLines	age	Nun59DaysPas
0	False	False	False	False	
1	False	False	False	False	
2	False	False	False	False	
3	False	False	False	False	
4	False	False	False	False	
...	
149995	False	False	False	False	
149996	False	False	False	False	
149997	False	False	False	False	
149998	False	False	False	False	
149999	False	False	False	False	

150000 rows × 12 columns

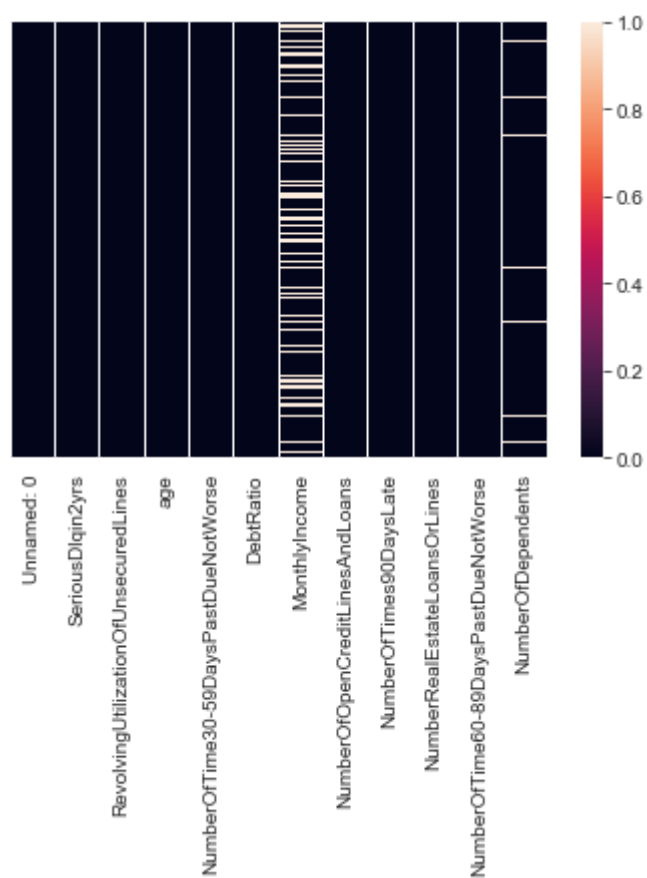


In [107]:

```
sns.heatmap(card.isnull(), yticklabels = False)
```

Out[107]:

<matplotlib.axes._subplots.AxesSubplot at 0xe8cfd88>



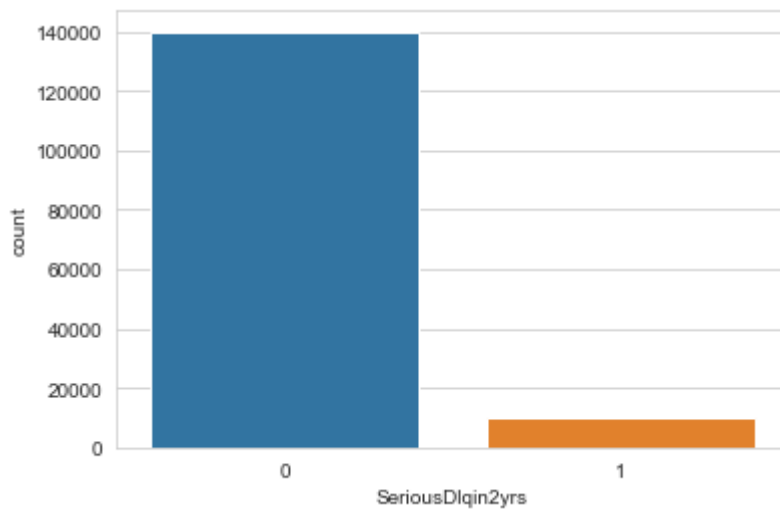
The graph shows null values are present in monthly income column and in number of dependents

In [108]:

```
sns.set_style('whitegrid')  
sns.countplot(x = 'SeriousDlqin2yrs', data = card)
```

Out[108]:

<matplotlib.axes._subplots.AxesSubplot at 0x1019b148>



Dependent Variable: 'seriousdlqin2yrs' with '1' (deliquent) and '0' (non-deliquent) as data points.

From the records, we have very less deliquent records, so, the model may be biased.

In [109]:

```
card.shape
```

Out[109]:

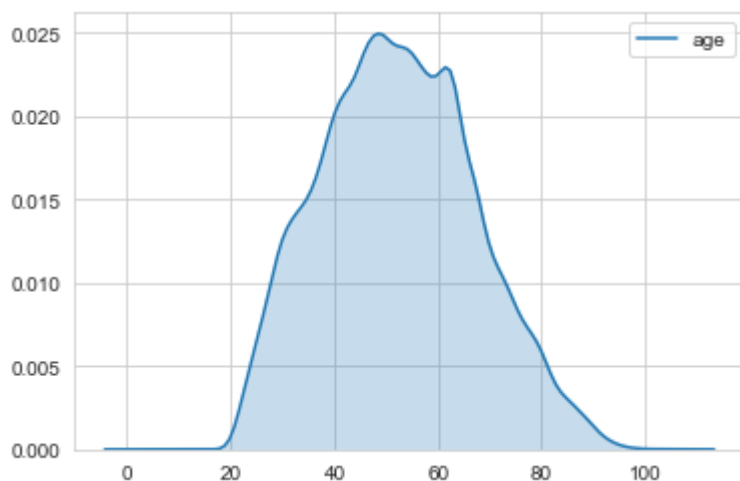
(150000, 12)

In [110]:

```
sns.kdeplot(card['age'], shade = True)
```

Out[110]:

<matplotlib.axes._subplots.AxesSubplot at 0xe842588>



In [111]:

```
card['RevolvingUtilizationOfUnsecuredLines'] = card['RevolvingUtilizationOfUnsecuredLines'] * 100
```

In [112]:

```
card['RevolvingUtilizationOfUnsecuredLines'] = card['RevolvingUtilizationOfUnsecuredLines'].apply(np.ceil)
```

In [113]:

```
card.head()
```

Out[113]:

	Unnamed: 0	SeriousDlqin2yrs	RevolvingUtilizationOfUnsecuredLines	age	NumberOfTimes59DaysPastDue
0	1	1	77.0	45	
1	2	0	96.0	40	
2	3	0	66.0	38	
3	4	0	24.0	30	
4	5	0	91.0	49	

In [114]:

```
card[card['RevolvingUtilizationOfUnsecuredLines']>100.0]
```

Out[114]:

	Unnamed: 0	SeriousDlqin2yrs	RevolvingUtilizationOfUnsecuredLines	age	NumberOfTimes59DaysPastDue
162	163	1	105.0	47	
191	192	0	110.0	53	
226	227	1	196.0	38	
251	252	1	105.0	58	
293	294	0	234000.0	45	
...
149939	149940	0	105.0	26	
149955	149956	1	114.0	41	
149962	149963	0	101.0	48	
149964	149965	0	102.0	63	
149973	149974	0	103.0	44	

3321 rows × 12 columns

There are some records >100.0

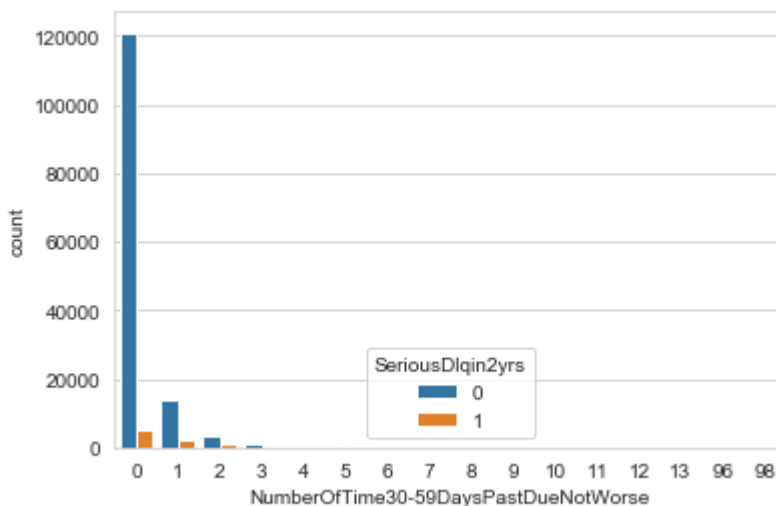
In [115]:

```
sns.countplot(card['NumberOfTime30-59DaysPastDueNotWorse'], hue = card['SeriousDlqin2yrs'])
card['NumberOfTime30-59DaysPastDueNotWorse'].value_counts()
```

Out[115]:

```
0    126018
1     16033
2      4598
3      1754
4       747
5       342
98      264
6       140
7        54
8        25
9         12
96         5
10         4
12         2
13         1
11         1
```

Name: NumberOfTime30-59DaysPastDueNotWorse, dtype: int64



In [116]:

```
card.columns
```

Out[116]:

```
Index(['Unnamed: 0', 'SeriousDlqin2yrs',
      'RevolvingUtilizationOfUnsecuredLines', 'age',
      'NumberOfTime30-59DaysPastDueNotWorse', 'DebtRatio', 'MonthlyIncome',
      'NumberOfOpenCreditLinesAndLoans', 'NumberOfTimes90DaysLate',
      'NumberRealEstateLoansOrLines', 'NumberOfTime60-89DaysPastDueNotWorse',
      'NumberOfDependents'],
      dtype='object')
```

In [118]:

```
monthly_inc_median = card['MonthlyIncome'].median()
```

In [119]:

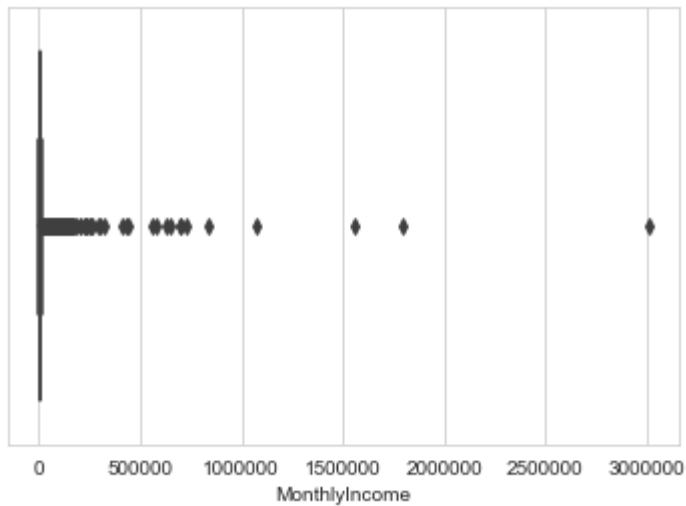
```
card['MonthlyIncome'] = card['MonthlyIncome'].fillna(monthly_inc_median)
```

In [121]:

```
sns.boxplot(card['MonthlyIncome'])
```

Out[121]:

<matplotlib.axes._subplots.AxesSubplot at 0xe4a3d08>

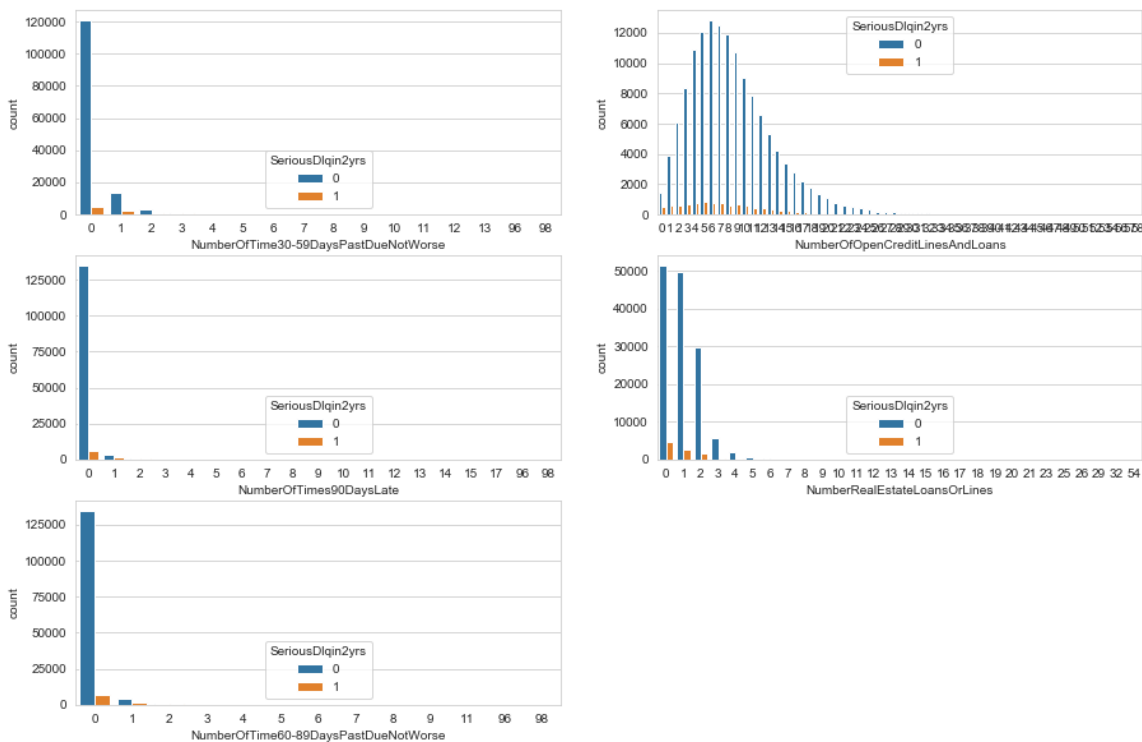


In [126]:

```
#creating subplot
f = plt.figure(figsize=(15,10))
a =f.add_subplot(3,2,1)
sns.countplot(x = card[ 'NumberOfTime30-59DaysPastDueNotWorse'], hue = card[ 'SeriousDlqin2yrs'])
plt.subplot(3,2,2)
sns.countplot(x = card[ 'NumberOfOpenCreditLinesAndLoans'], hue = card[ 'SeriousDlqin2yrs'])
plt.subplot(3,2,3)
sns.countplot(x = card[ 'NumberOfTimes90DaysLate'], hue = card[ 'SeriousDlqin2yrs'])
plt.subplot(3,2,4)
sns.countplot(x = card[ 'NumberRealEstateLoansOrLines'], hue = card[ 'SeriousDlqin2yrs'])
plt.subplot(3,2,5)
sns.countplot(x = card[ 'NumberOfTime60-89DaysPastDueNotWorse'], hue = card[ 'SeriousDlqin2yrs'])
```

Out[126]:

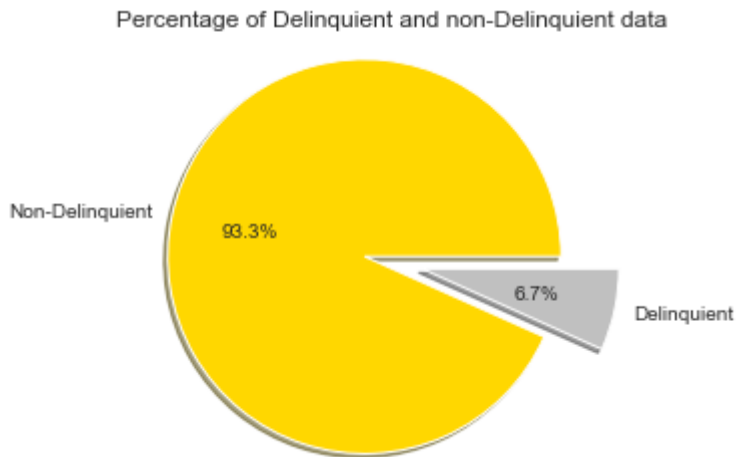
<matplotlib.axes._subplots.AxesSubplot at 0x118a1688>



In [150]:

```
labels = ['Non-Delinquent', 'Delinquent']
sizes = []
sizes.append(list(card['SeriousDlqin2yrs'].value_counts())[0])
sizes.append(list(card['SeriousDlqin2yrs'].value_counts())[1])
colors = ['gold', 'silver']

plt.pie(sizes, explode=(0.3,0), labels=labels, colors=colors, autopct='%1.1f%%', shadow
= True)
plt.title('Percentage of Delinquent and non-Delinquent data')
plt.axis('equal')
plt.show()
```



In []:

In []: