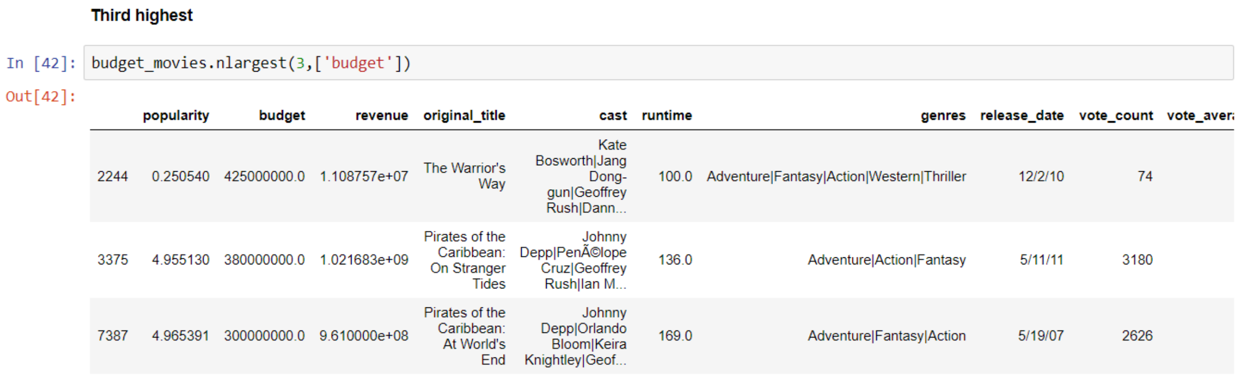
MINOR PROJECT (ML-MINOR-NOV)

ARUN PRAKASH

After cleaning the data and doing EDA, I have answered four questions as shown below.

1) Which are the movies with the third-lowest and third-highest budget?

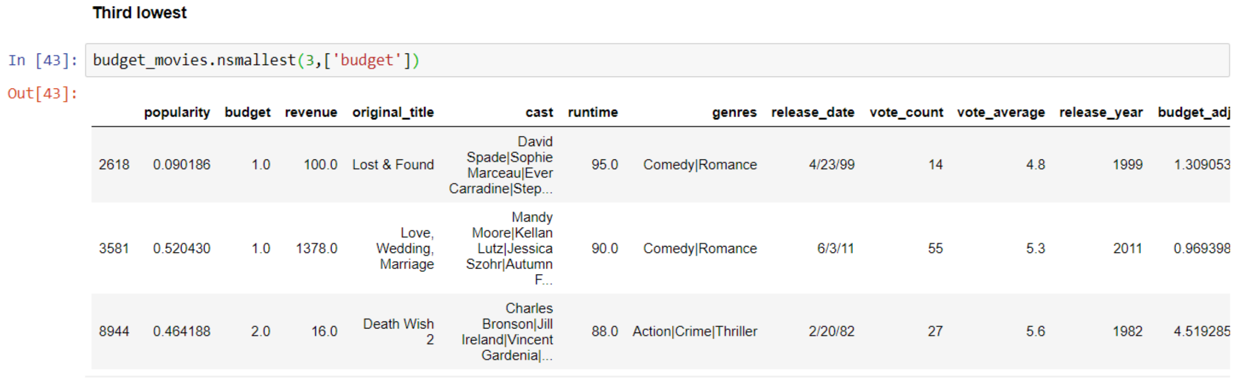


Third highest

```
In [42]: budget_movies.nlargest(3,['budget'])
Out[42]:
```

	popularity	budget	revenue	original_title	cast	runtime	genres	release_date	vote_count	vote_aver
2244	0.250540	425000000.0	1.108757e+07	The Warrior's Way	Kate Bosworth|Jang Dong-gun|Geoffrey Rush|Dann...	100.0	Adventure|Fantasy|Action|Western|Thriller	12/2/10	74	
3375	4.955130	380000000.0	1.021683e+09	Pirates of the Caribbean: On Stranger Tides	Johnny Depp|PenÃ©lope Cruz|Geoffrey Rush|Ian M...	136.0	Adventure|Action|Fantasy	5/11/11	3180	
7387	4.965391	300000000.0	9.610000e+08	Pirates of the Caribbean: At World's End	Johnny Depp|Orlando Bloom|Keira Knightley|Geof...	169.0	Adventure|Fantasy|Action	5/19/07	2626	

From the above output we can see that "Pirates of the Caribbean: At World's End" has the third highest budget.



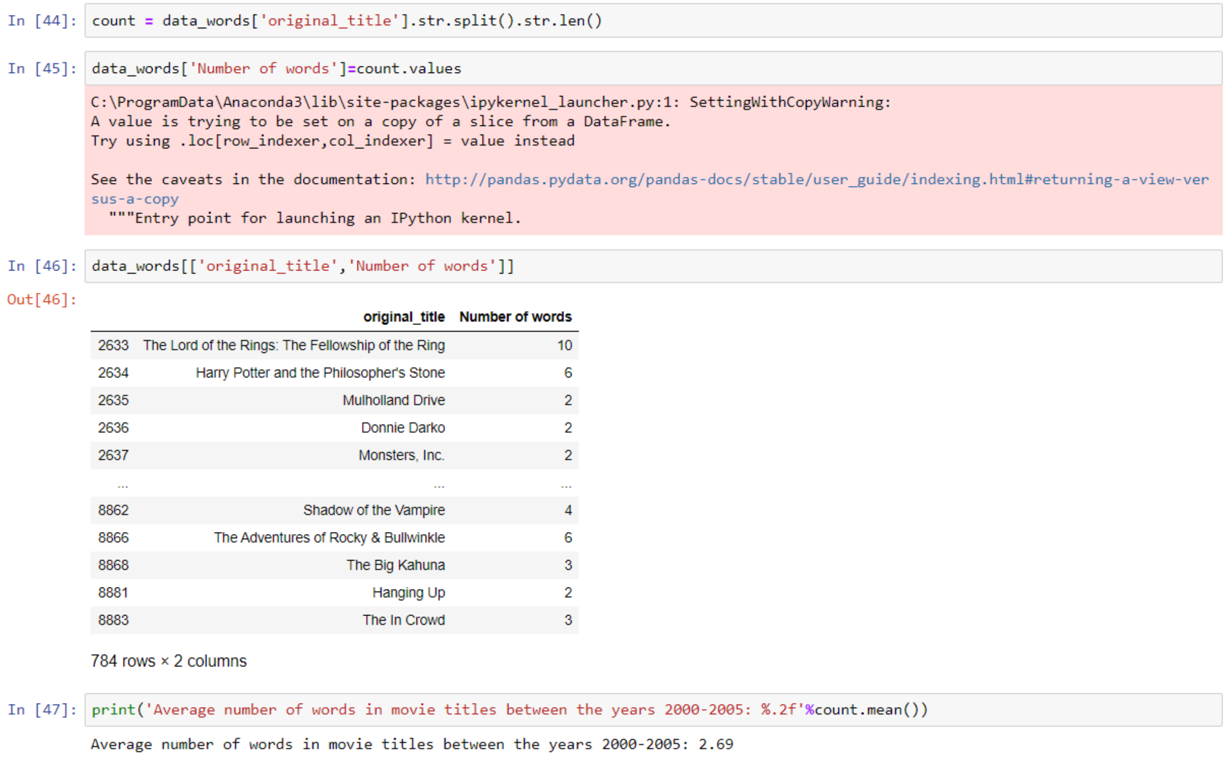
Third lowest

```
In [43]: budget_movies.nsmallest(3,['budget'])
Out[43]:
```

	popularity	budget	revenue	original_title	cast	runtime	genres	release_date	vote_count	vote_average	release_year	budget_adj
2618	0.090186	1.0	100.0	Lost & Found	David Spade|Sophie Marceau|Ever Carradine|Step...	95.0	Comedy|Romance	4/23/99	14	4.8	1999	1.309053
3581	0.520430	1.0	1378.0	Love, Wedding, Marriage	Mandy Moore|Kellan Lutz|Jessica Szohr|Autumn F...	90.0	Comedy|Romance	6/3/11	55	5.3	2011	0.969398
8944	0.464188	2.0	16.0	Death Wish 2	Charles Bronson|Jill Ireland|Vincent Gardenia|...	88.0	Action|Crime|Thriller	2/20/82	27	5.6	1982	4.519285

From the above output we can see that "Death Wish 2" has the third lowest budget.

2) What is the average number of words in movie titles between the years 2000-2005?



```
In [44]: count = data_words['original_title'].str.split().str.len()
```

```
In [45]: data_words['Number of words']=count.values
```

```
C:\ProgramData\Anaconda3\lib\site-packages\ipykernel_launcher.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-ver
sus-a-copy
  """Entry point for launching an IPython kernel.
```

```
In [46]: data_words[['original_title','Number of words']]
```

Out[46]:

	original_title	Number of words
2633	The Lord of the Rings: The Fellowship of the Ring	10
2634	Harry Potter and the Philosopher's Stone	6
2635	Mulholland Drive	2
2636	Donnie Darko	2
2637	Monsters, Inc.	2
...	...	...
8862	Shadow of the Vampire	4
8866	The Adventures of Rocky & Bullwinkle	6
8868	The Big Kahuna	3
8881	Hanging Up	2
8883	The In Crowd	3

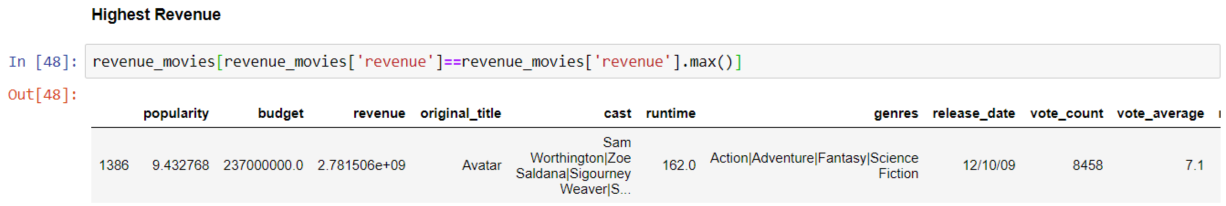
784 rows × 2 columns

```
In [47]: print('Average number of words in movie titles between the years 2000-2005: %.2f'%count.mean())
```

```
Average number of words in movie titles between the years 2000-2005: 2.69
```

From the above output we can see that the average number of words in movie titles between the years 2000-2005 is between 2 to 3 words.

4) Which are the movies with the most and least earned revenue?



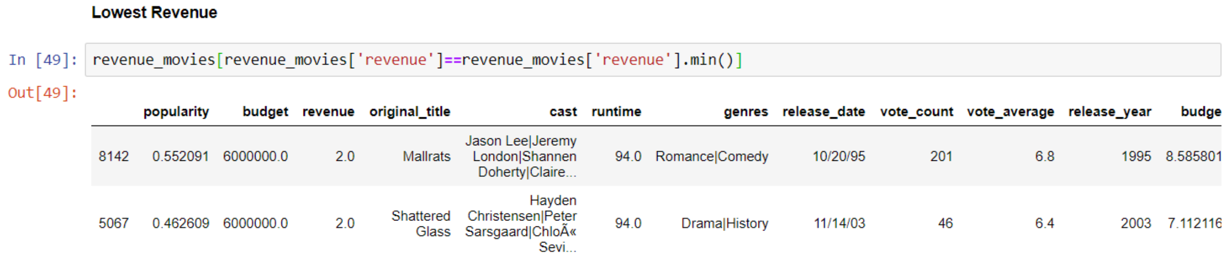
Highest Revenue

```
In [48]: revenue_movies[revenue_movies['revenue']==revenue_movies['revenue'].max()]
```

Out[48]:

	popularity	budget	revenue	original_title	cast	runtime	genres	release_date	vote_count	vote_average
1386	9.432768	237000000.0	2.781506e+09	Avatar	Sam Worthington|Zoe Saldana|Sigourney Weaver|S...	162.0	Action|Adventure|Fantasy|Science Fiction	12/10/09	8458	7.1

From the above output we can see that the revenue of "Avatar" is the highest.



Lowest Revenue

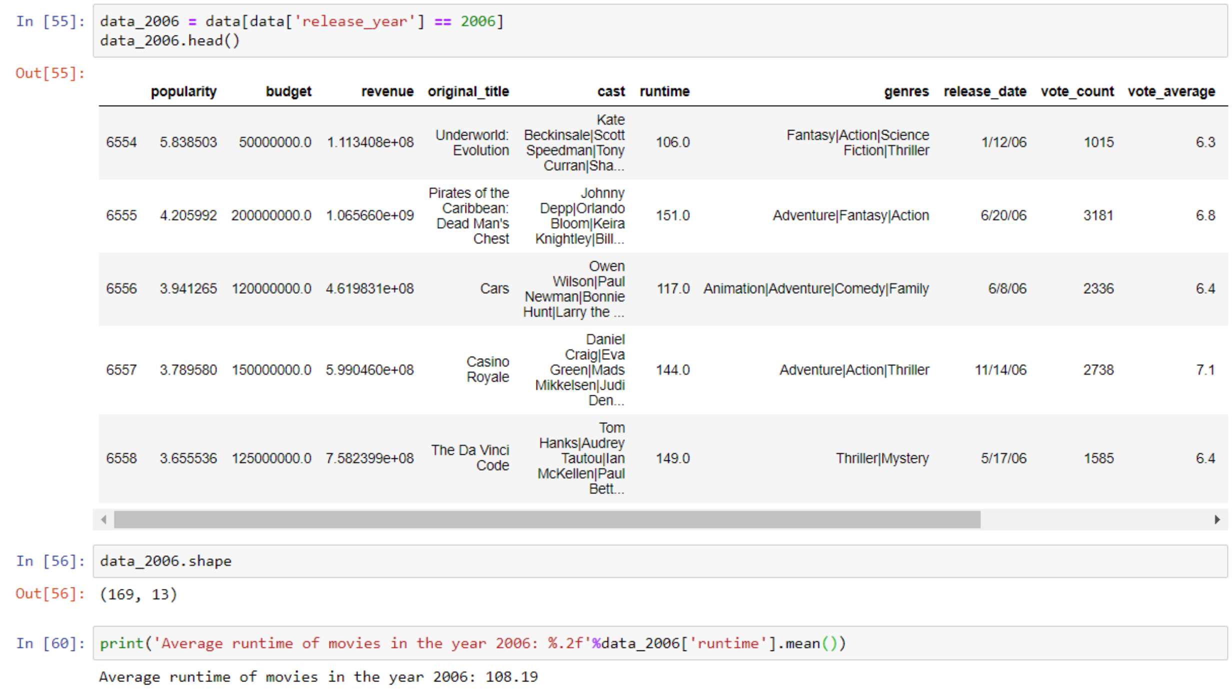
```
In [49]: revenue_movies[revenue_movies['revenue']==revenue_movies['revenue'].min()]
```

Out[49]:

	popularity	budget	revenue	original_title	cast	runtime	genres	release_date	vote_count	vote_average	release_year	budge
8142	0.552091	6000000.0	2.0	Mallrats	Jason Lee|Jeremy London|Shannen Doherty|Claire...	94.0	Romance|Comedy	10/20/95	201	6.8	1995	8.585801
5067	0.462609	6000000.0	2.0	Shattered Glass	Hayden Christensen|Peter Sarsgaard|ChloÃ« Sevi...	94.0	Drama|History	11/14/03	46	6.4	2003	7.112116

From the above output we can see that "Mallrats" and "Shattered Glass" got the least revenue.

5) What is the average runtime of movies in the year 2006?



```
In [55]: data_2006 = data[data['release_year'] == 2006]
         data_2006.head()
```

Out[55]:

	popularity	budget	revenue	original_title	cast	runtime	genres	release_date	vote_count	vote_average
6554	5.838503	50000000.0	1.113408e+08	Underworld: Evolution	Kate Beckinsale|Scott Speedman|Tony Curran|Sha...	106.0	Fantasy|Action|Science Fiction|Thriller	1/12/06	1015	6.3
6555	4.205992	200000000.0	1.065660e+09	Pirates of the Caribbean: Dead Man's Chest	Johnny Depp|Orlando Bloom|Keira Knightley|Bill...	151.0	Adventure|Fantasy|Action	6/20/06	3181	6.8
6556	3.941265	120000000.0	4.619831e+08	Cars	Owen Wilson|Paul Newman|Bonnie Hunt|Larry the ...	117.0	Animation|Adventure|Comedy|Family	6/8/06	2336	6.4
6557	3.789580	150000000.0	5.990460e+08	Casino Royale	Daniel Craig|Eva Green|Mads Mikkelsen|Judi Den...	144.0	Adventure|Action|Thriller	11/14/06	2738	7.1
6558	3.655536	125000000.0	7.582399e+08	The Da Vinci Code	Tom Hanks|Audrey Tautou|Ian McKellen|Paul Bett...	149.0	Thriller|Mystery	5/17/06	1585	6.4

```
In [56]: data_2006.shape

Out[56]: (169, 13)

In [60]: print('Average runtime of movies in the year 2006: %.2f'%data_2006['runtime'].mean())

         Average runtime of movies in the year 2006: 108.19
```

From the above output we can see that the average runtime of movies in the year 2006 is 108.49 minutes.