In [2]:
```
import pandas as pd
import seaborn as sns
import numpy as np
import matplotlib.pyplot as plt
import warnings
warnings.filterwarnings ('ignore')
import requests

df = pd.read_csv (r"C:\Users\nette802\Desktop\Data Science\data science.csv")
```

In [3]:
```
df
```

Out[3]:

	Unnamed: 0	work_year	experience_level	employment_type	job_title	salary	salary_currency
0	0	2020	MI	FT	Data Scientist	70000	EUR
1	1	2020	SE	FT	Machine Learning Scientist	260000	USD
2	2	2020	SE	FT	Big Data Engineer	85000	GBP
3	3	2020	MI	FT	Product Data Analyst	20000	USD
4	4	2020	SE	FT	Machine Learning Engineer	150000	USD
...	...	...	...	...	...	...	...
602	602	2022	SE	FT	Data Engineer	154000	USD
603	603	2022	SE	FT	Data Engineer	126000	USD
604	604	2022	SE	FT	Data Analyst	129000	USD
605	605	2022	SE	FT	Data Analyst	150000	USD
606	606	2022	MI	FT	AI Scientist	200000	USD

607 rows × 12 columns

In [15]:
```
# select two columns
df[['salary','job_title']]
```

Out[15]:

	salary	job_title
0	70000	Data Scientist
1	260000	Machine Learning Scientist
2	85000	Big Data Engineer
3	20000	Product Data Analyst
4	150000	Machine Learning Engineer
...	...	...
602	154000	Data Engineer
603	126000	Data Engineer
604	129000	Data Analyst
605	150000	Data Analyst
606	200000	AI Scientist

607 rows × 2 columns

In [14]:
```
#FINDING MAX AND MIN OF SALARY
p=df['salary'].max()
q=df['salary'].min()
```

In [9]:
```
print (p)
```

```
30400000
```

In [10]:
```
print (q)
```

```
4000
```

In [17]:
```
df.sort_values(by = ['salary', 'job_title'], axis=0, ascending=[False, True], inpl
               kind='quicksort', na_position='first', ignore_index=True, key=None)
```

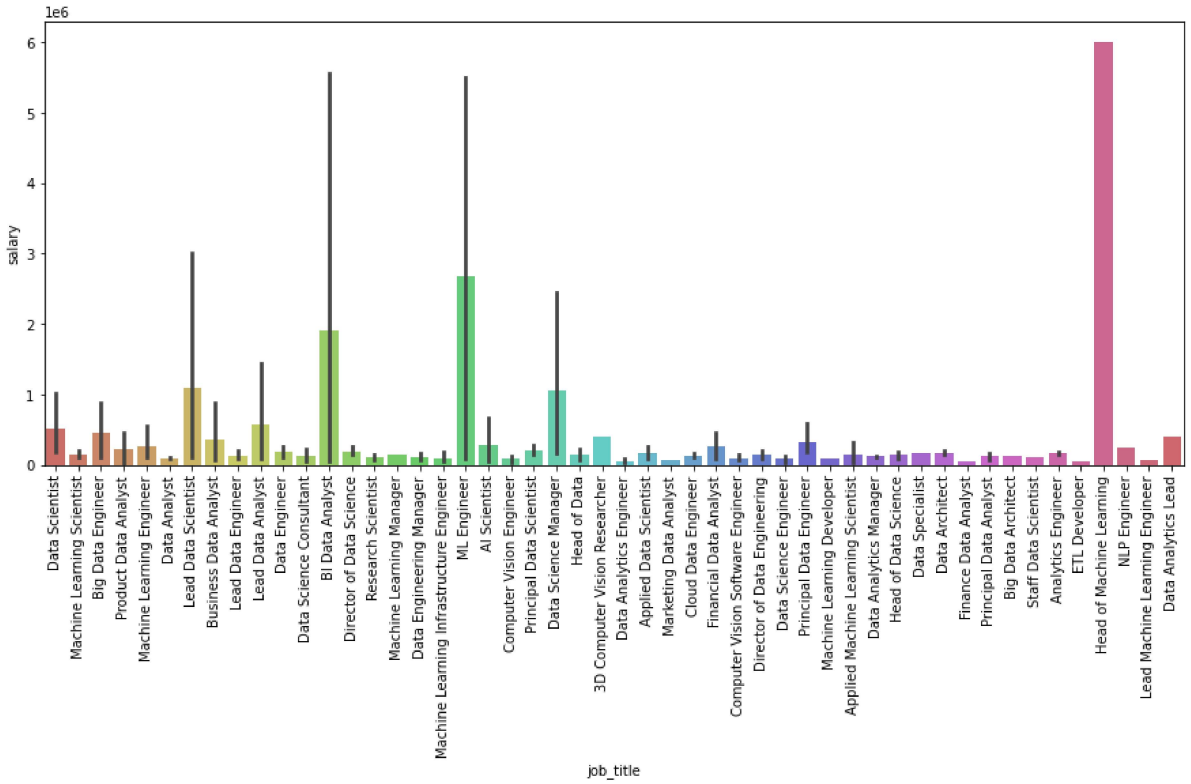
Out[17]:

	Unnamed: 0	work_year	experience_level	employment_type	job_title	salary	salary_curren
0	177	2021	MI	FT	Data Scientist	30400000	C
1	102	2021	MI	FT	BI Data Analyst	11000000	H
2	7	2020	MI	FT	Data Scientist	11000000	H
3	137	2021	MI	FT	ML Engineer	8500000	J
4	285	2021	SE	FT	Data Science Manager	7000000	II
...	...	...	...	...	...	...	
602	196	2021	EN	FT	BI Data Analyst	9272	U
603	124	2021	EN	PT	Data Analyst	8760	E
604	15	2020	MI	FT	Data Analyst	8000	U
605	185	2021	MI	FT	Data Engineer	4000	U
606	238	2021	EN	FT	Data Scientist	4000	U

607 rows × 12 columns

In [21]:
```
plt.figure(figsize=(15,6))
sns.barplot(y = 'salary', x = 'job_title', data= df, palette = 'hls' )
plt.xticks(rotation = 90)
plt.show
```

Out[21]:
```
<function matplotlib.pyplot.show(close=None, block=None)>
```



In [23]:
```
df.sort_values(by = ['salary', 'salary_currency'], axis=0, ascending=[False, True],
               kind='quicksort', na_position='first', ignore_index=True, key=None)
```

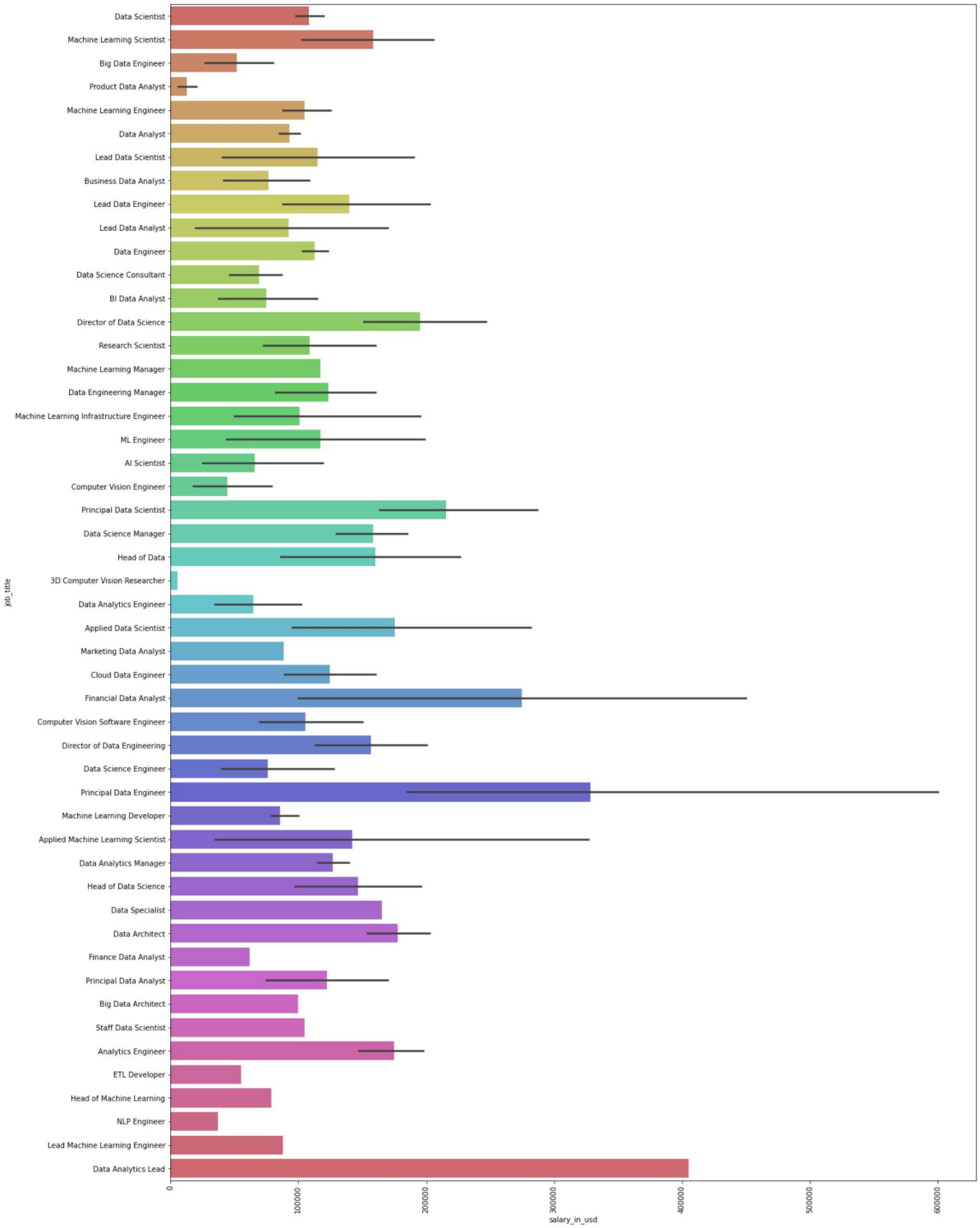
Out[23]:

	Unnamed: 0	work_year	experience_level	employment_type	job_title	salary	salary_curren
0	177	2021	MI	FT	Data Scientist	30400000	C
1	7	2020	MI	FT	Data Scientist	11000000	H
2	102	2021	MI	FT	BI Data Analyst	11000000	H
3	137	2021	MI	FT	ML Engineer	8500000	J
4	285	2021	SE	FT	Data Science Manager	7000000	II
...	...	...	...	...	...	...	
602	196	2021	EN	FT	BI Data Analyst	9272	U
603	124	2021	EN	PT	Data Analyst	8760	E
604	15	2020	MI	FT	Data Analyst	8000	U
605	185	2021	MI	FT	Data Engineer	4000	U
606	238	2021	EN	FT	Data Scientist	4000	U

607 rows × 12 columns

In [32]:
```
plt.figure(figsize=(20,30))
sns.barplot(y = 'job_title', x = 'salary_in_usd', data= df, palette = 'hls' )
plt.xticks(rotation = 90)
plt.show
```

Out[32]:
```
<function matplotlib.pyplot.show(close=None, block=None)>
```



In [33]:
```
df.sort_values(by = ['remote_ratio', 'job_title'], axis=0, ascending=[False, True],
               kind='quicksort', na_position='first', ignore_index=True, key=None)
```

Out[33]:

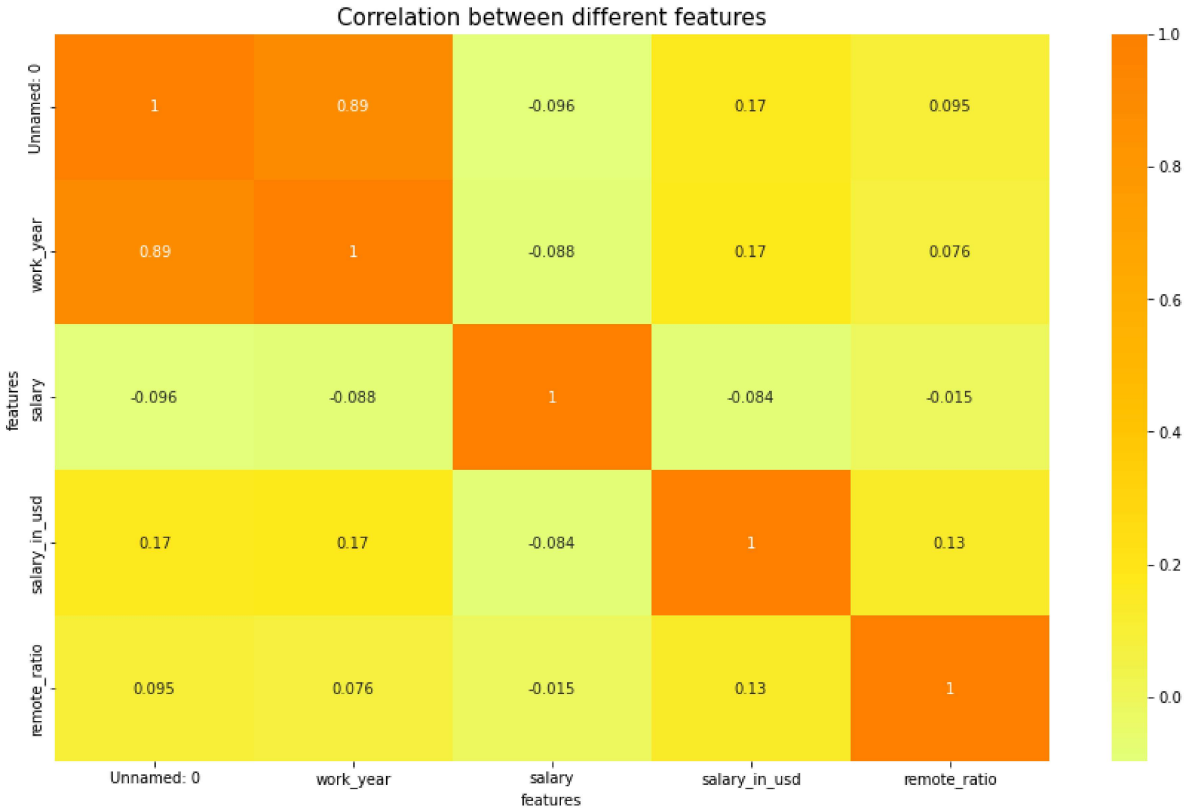
	Unnamed: 0	work_year	experience_level	employment_type	job_title	salary	salary_currenc
0	96	2021	EN	PT	AI Scientist	12000	US
1	113	2021	EN	PT	AI Scientist	12000	US
2	244	2021	EN	FT	AI Scientist	1335000	IN
3	277	2021	SE	FT	AI Scientist	55000	US
4	606	2022	MI	FT	AI Scientist	200000	US
...	...	...	...	...	...	...	
602	138	2021	SE	FT	Principal Data Scientist	220000	US
603	3	2020	MI	FT	Product Data Analyst	20000	US
604	33	2020	MI	FT	Research Scientist	450000	US
605	281	2021	EN	FT	Research Scientist	100000	US
606	507	2022	MI	FT	Research Scientist	59000	EU

607 rows × 12 columns

In [49]:

```
fig, ax = plt.subplots(1,1, figsize=(15,9))
sns.heatmap (df.corr(), annot=True,cmap= 'Wistia')
ax.set(xlabel= 'features')
ax.set(ylabel='features')

plt.title('Correlation between different features', fontsize=15,c='black')
plt.show()
```



In [53]:
```
p=df['salary_in_usd'].max()
q=df['salary_in_usd'].min()
```

In [54]:
```
print (p)
```

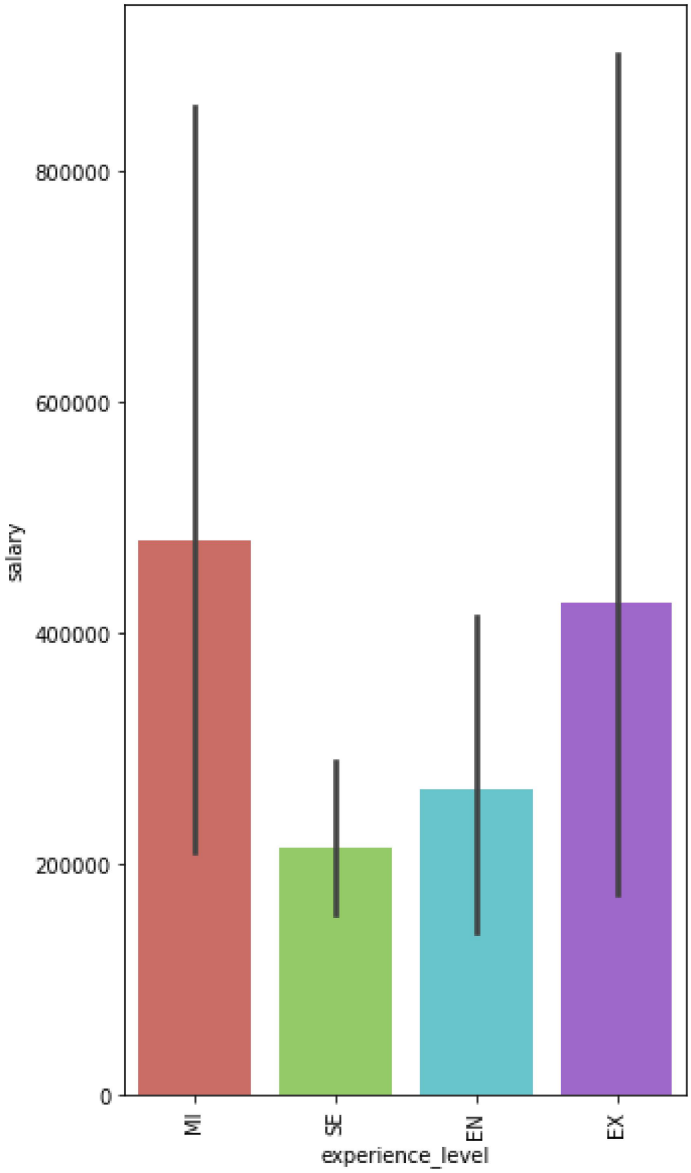
```
600000
```

In [55]:
```
print (q)
```

```
2859
```

In [63]:
```
plt.figure(figsize=(5,10))
sns.barplot(y = 'salary', x = 'experience_level'  , data= df, palette = 'hls' )
plt.xticks(rotation = 90)
plt.show
```

Out[63]:
```
<function matplotlib.pyplot.show(close=None, block=None)>
```


800000
600000
400000
200000
0
salary
MI
SE
EN
EX
experience_level