In [15]:
```
import pandas as pd
import matplotlib.pyplot as plt
stem = pd.read_csv("C:/Users/Abhinav/Desktop/DataAnalysisCourseMaterials/DataAnalysis/
stem
```

Out[15]:

	timestamp	company	level	title	totalyearlycompensation	location	yearsofexperienc
0	6/7/2017 11:33:27	Oracle	L3	Product Manager	127000	Redwood City, CA	1
1	6/10/2017 17:11:29	eBay	SE 2	Software Engineer	100000	San Francisco, CA	5
2	6/11/2017 14:53:57	Amazon	L7	Product Manager	310000	Seattle, WA	8
3	6/17/2017 0:23:14	Apple	M1	Software Engineering Manager	372000	Sunnyvale, CA	7
4	6/20/2017 10:58:51	Microsoft	60	Software Engineer	157000	Mountain View, CA	5
...	...	...	...	...	...	...	
62637	9/9/2018 11:52:32	Google	T4	Software Engineer	327000	Seattle, WA	10
62638	9/13/2018 8:23:32	Microsoft	62	Software Engineer	237000	Redmond, WA	2
62639	9/13/2018 14:35:59	MSFT	63	Software Engineer	220000	Seattle, WA	14
62640	9/16/2018 16:10:35	Salesforce	Lead MTS	Software Engineer	280000	San Francisco, CA	8
62641	1/29/2019 5:12:59	apple	ict3	Software Engineer	200000	Sunnyvale, CA	0

62642 rows × 29 columns

In [9]:
```
stem.shape
```

Out[9]:
```
(62642, 29)
```

In [29]:
```
STEM = stem.drop(labels=["Doctorate_Degree","Highschool","Some_College","Race_Asian","
STEM
```

Out[29]:

	company	level	title	totalyearlycompensation	location	yearsofexperience	yearsatco
0	Oracle	L3	Product Manager	127000	Redwood City, CA	1.5	
1	eBay	SE 2	Software Engineer	100000	San Francisco, CA	5.0	
2	Amazon	L7	Product Manager	310000	Seattle, WA	8.0	
3	Apple	M1	Software Engineering Manager	372000	Sunnyvale, CA	7.0	
4	Microsoft	60	Software Engineer	157000	Mountain View, CA	5.0	
...	...	...	...	...	...	...	
62637	Google	T4	Software Engineer	327000	Seattle, WA	10.0	
62638	Microsoft	62	Software Engineer	237000	Redmond, WA	2.0	
62639	MSFT	63	Software Engineer	220000	Seattle, WA	14.0	
62640	Salesforce	Lead MTS	Software Engineer	280000	San Francisco, CA	8.0	
62641	apple	ict3	Software Engineer	200000	Sunnyvale, CA	0.0	

62642 rows × 13 columns

In [32]:
```
STEM.shape
```

Out[32]:
```
(62642, 13)
```

In [37]:
```
maang_companies =['Facebook','Apple','Amazon','Netflix','Google']
STEM_Maang = STEM[STEM['company'].isin(maang_companies)]
STEM_Maang
```

Out[37]:

	company	level	title	totalyearlycompensation	location	yearsofexperience	yearsatco
2	Amazon	L7	Product Manager	310000	Seattle, WA	8.0	
3	Apple	M1	Software Engineering Manager	372000	Sunnyvale, CA	7.0	
14	Amazon	L6	Software Engineering Manager	287000	Seattle, WA	12.0	
15	Amazon	L5	Software Engineer	218000	Seattle, WA	10.0	
16	Facebook	E3	Software Engineer	168000	Menlo Park, CA	1.0	
...	...	...	...	...	...	...	
62621	Google	L4	Software Engineer	215000	Mountain View, CA	2.0	
62627	Google	T3	Software Engineer	169000	San Bruno, CA	0.0	
62631	Facebook	E5	Software Engineering Manager	421000	Menlo Park, CA	13.0	
62634	Amazon	L5	Software Engineer	213000	Seattle, WA	6.0	
62637	Google	T4	Software Engineer	327000	Seattle, WA	10.0	

17742 rows × 13 columns

In [56]:
```
summary = STEM_Maang.describe()
summary
```

Out[56]:

	totalyearlycompensation	yearsofexperience	yearsatcompany	basesalary	stockgrantvalue
count	1.774200e+04	17742.000000	17742.000000	17742.000000	1.774200e+04
mean	2.703695e+05	7.218752	2.254292	147347.394995	8.509919e+04
std	1.584696e+05	5.765610	2.596207	67247.638121	1.001946e+05
min	1.000000e+04	0.000000	0.000000	0.000000	0.000000e+00
25%	1.790000e+05	3.000000	0.000000	122000.000000	2.500000e+04
50%	2.360000e+05	6.000000	1.800000	150000.000000	5.600000e+04
75%	3.200000e+05	10.000000	3.000000	170000.000000	1.100000e+05
max	4.980000e+06	42.000000	27.000000	893000.000000	2.800000e+06

In [44]:
```
SFO_data = STEM_Maang [STEM_Maang['location'] == 'San Francisco, CA']
SFO_data
```

Out[44]:

	company	level	title	totalyearlycompensation	location	yearsofexperience	yearsatcor
40	Amazon	L5	Software Engineer	250000	San Francisco, CA	12.0	
78	Apple	ICT4	Product Manager	305000	San Francisco, CA	7.0	
93	Apple	M1	Software Engineering Manager	385000	San Francisco, CA	7.0	
107	Amazon	L5	Software Engineer	250000	San Francisco, CA	10.0	
113	Amazon	L6	Software Engineering Manager	500000	San Francisco, CA	15.0	
...	...	...	...	...	...	...	
62398	Amazon	L4	Software Engineer	145000	San Francisco, CA	1.0	
62494	Amazon	L5	Software Engineer	216000	San Francisco, CA	17.0	
62499	Facebook	E4	Software Engineer	410000	San Francisco, CA	6.0	
62580	Google	L5	Product Manager	298000	San Francisco, CA	11.0	
62593	Amazon	L5	Software Engineer	311000	San Francisco, CA	10.0	

1127 rows × 13 columns

In [45]:
```
Product_Manager = SFO_data [SFO_data['title'] == 'Product Manager']
Product_Manager
```

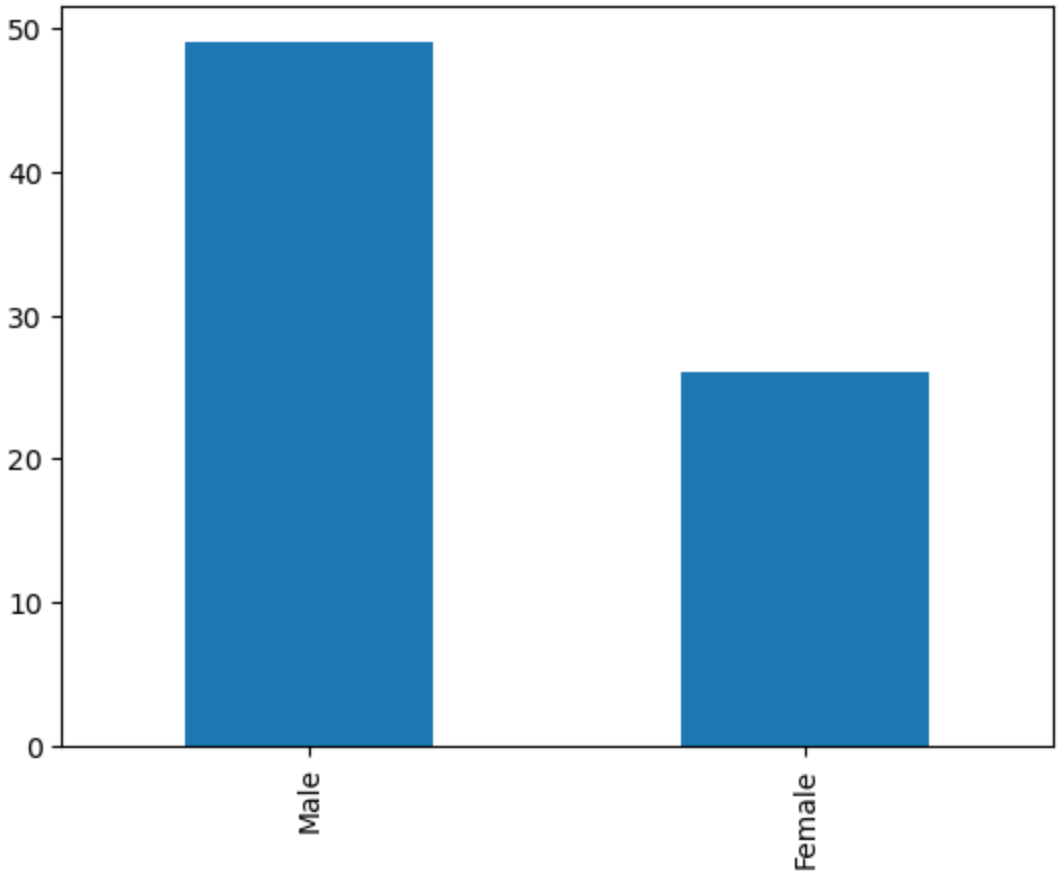
Out[45]:

	company	level	title	totalyearlycompensation	location	yearsofexperience	yearsatco
78	Apple	ICT4	Product Manager	305000	San Francisco, CA	7.0	
156	Google	L5	Product Manager	350000	San Francisco, CA	5.0	
209	Amazon	L7	Product Manager	430000	San Francisco, CA	20.0	
1077	Facebook	L7 Product Manager	Product Manager	600000	San Francisco, CA	12.0	
1078	Facebook	L7 Product Manager	Product Manager	600000	San Francisco, CA	12.0	
...	...	...	...	...	...	...	
60699	Amazon	L8	Product Manager	850000	San Francisco, CA	12.0	
61022	Amazon	Senior Product Manager	Product Manager	305000	San Francisco, CA	15.0	
61982	Google	L10	Product Manager	4500000	San Francisco, CA	20.0	
62336	Google	7	Product Manager	530000	San Francisco, CA	15.0	
62580	Google	L5	Product Manager	298000	San Francisco, CA	11.0	

123 rows × 13 columns

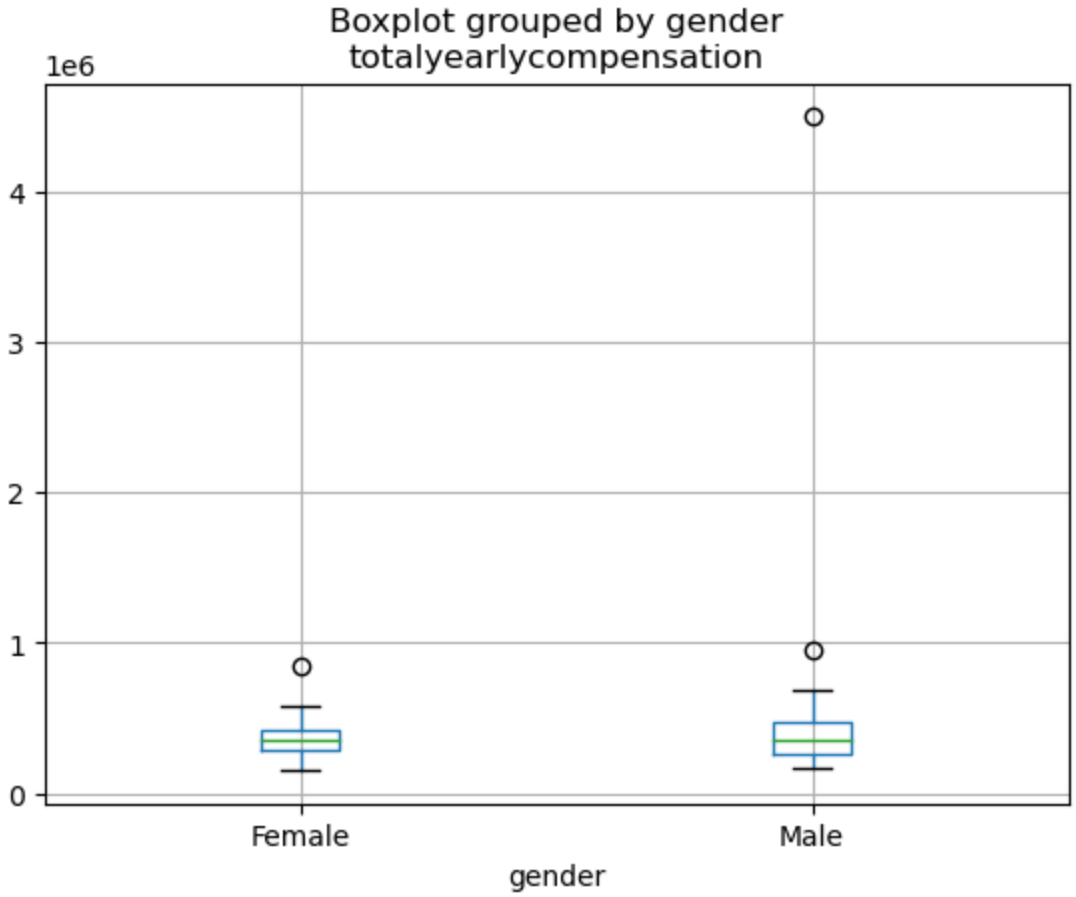
In [62]:
```
gender_counts = Product_Manager['gender'].value_counts()
gender_counts.plot(kind='bar')
```

Out[62]:
```
<Axes: >
```



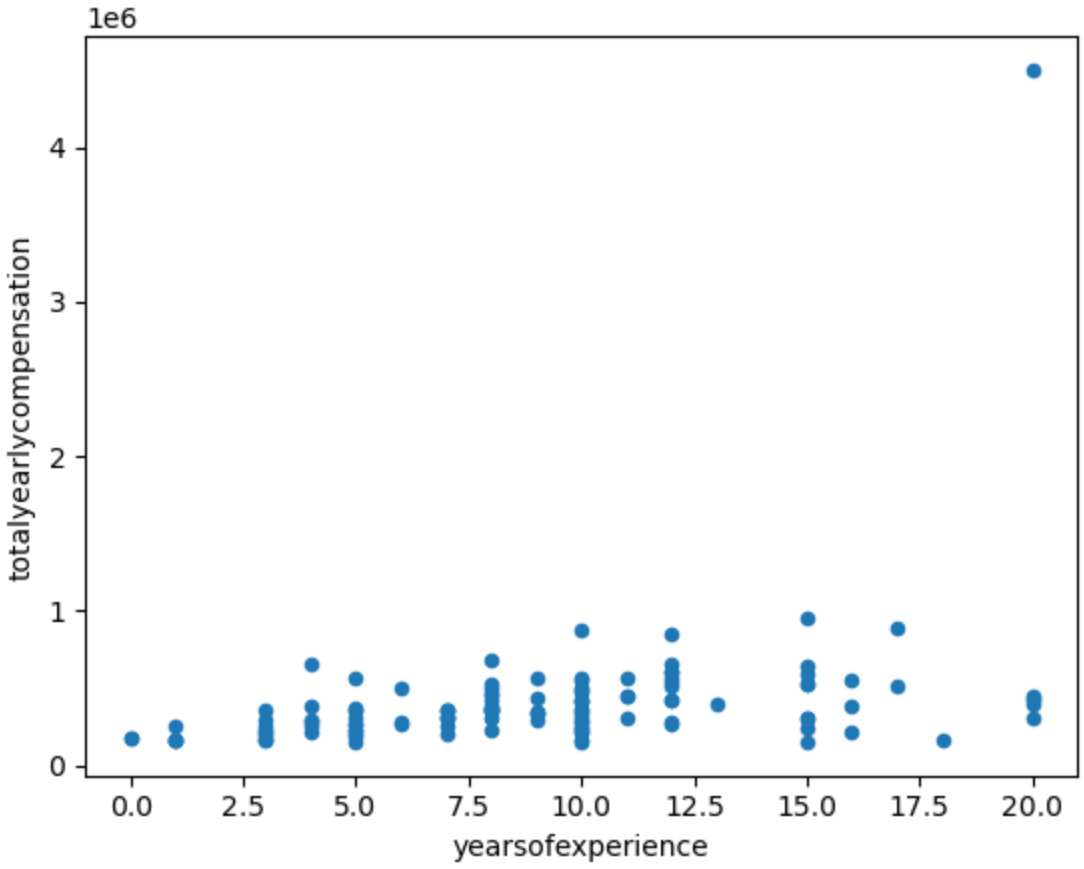
In [63]:
```
Product_Manager.boxplot(column='totalyearlycompensation', by='gender')
```

Out[63]:
```
<Axes: title={'center': 'totalyearlycompensation'}, xlabel='gender'>
```



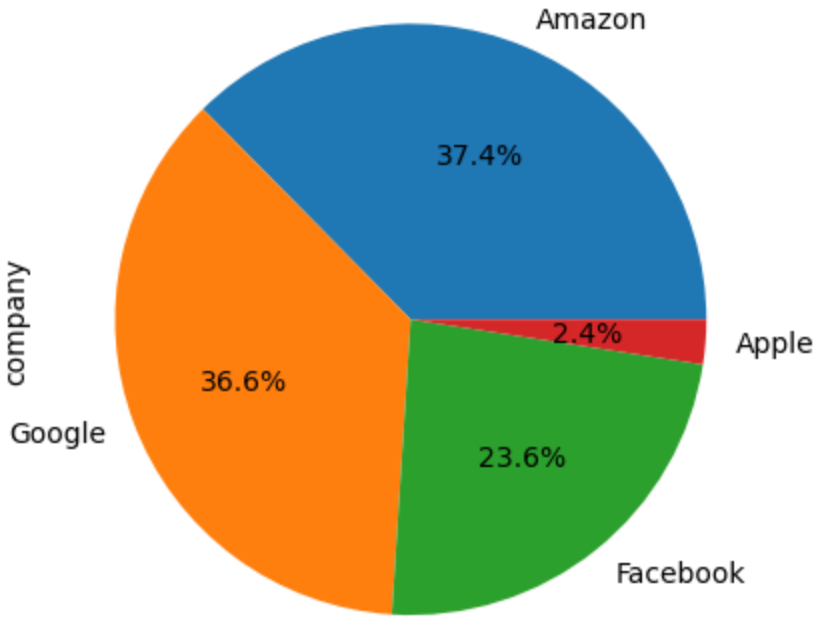
In [64]:
```
Product_Manager.plot.scatter(x='yearsofexperience', y='totalyearlycompensation')
```

Out[64]:
```
<Axes: xlabel='yearsofexperience', ylabel='totalyearlycompensation'>
```



In [65]:
```
company_distribution = Product_Manager['company'].value_counts()
company_distribution.plot(kind='pie', autopct='%1.1f%%')
```

Out[65]:
```
<Axes: ylabel='company'>
```



In [46]:
```
Experience_Years = Product_Manager[Product_Manager['yearsofexperience'] == 0]
Experience_Years
```

Out[46]:

	company	level	title	totalyearlycompensation	location	yearsofexperience	yearsatcompa
20991	Google	L3	Product Manager	170000	San Francisco, CA	0.0	
33711	Google	L3	Product Manager	175000	San Francisco, CA	0.0	

In [66]:
```
Experience_Years = Product_Manager[(Product_Manager['yearsofexperience'] >= 0) & (Proc
plt.figure(figsize=(10, 6))
plt.hist(Experience_Years['yearsofexperience'], bins=16, color='skyblue', edgecolor='b
plt.xlabel('Years of Experience')
plt.ylabel('Number of Product Managers')
plt.title('Distribution of Product Managers by Years of Experience (0-15 years)')
plt.grid(axis='y', linestyle='--', alpha=0.7)

# Show the plot
plt.tight_layout()
plt.show()
```

In [ ]: