1. **Age appears to be a significant factor in this analysis. Calculate the patients' ages based on their dates of birth.**

2. **You should also visualize the distribution of costs using a histogram, box and whisker plot, and swarm plot.**

The first screenshot shows an Excel spreadsheet with the "export_dataframe" sheet and a scatter chart titled "Age". The table columns visible:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Customer	year	month	Month_nu	day	Date	children	charges	Age	Hospital ti	City tier	State ID	BMI	HBA1C	Heart Issu	Any Trans	Cancer hi:
2	Id2335	1992	Sep	9	9	09-09-1992	0	563.84	30	2	3	R1013	17.58	4.51	0	0	0
3	Id2334	1992	Nov	11	30	30-11-1992	0	570.62	30								1
4	Id2333	1993	Jun	6	30	30-06-1993	0	600	29								1
5	Id2332	1992	Sep	9	13	13-09-1992	0	604.54	30								0
6	Id2331	1998	Jul	7	27	27-07-1998	0	637.26	24								0
7	Id2330	2001	Nov	11	20	20-11-2001	0	646.14	21								0
8	Id2329	1993	Jun	6	1	01-06-1993	0	650	29								1
9	Id2328	1995	Jul	7	4	04-07-1995	0	650	27								0
10	Id2327	2002	Nov	11	29	29-11-2002	0	668	20								0
11	Id2326	1997	Nov	11	9	09-11-1997	0	670	25								1
12	Id2325	2001	Sep	9	12	12-09-2001	0	687.54	21								0
13	Id2324	1999	Dec	12	26	26-12-1999	0	700	23								0
14	Id2323	1999	Dec	12	14	14-12-1999	0	722.99	23								0
15	Id2322	2002		1	19	19-01-2002	0	750	21								0
16	Id2321	1993	Aug	8	9	09-08-1993	0	760	29								1
17	Id2320	1996	Oct	10	22	22-10-1996	0	760	26								0
18	Id2319	1993	Jun	6	28	28-06-1993	0	770	29	3	3	R1013	17.08	5.73	0	0	1
19	Id2318	1996		1	18	18-01-1996	0	770.38	27	3	2	R1012	18.82	5.51	1	0	0
20	Id2317	1995	Dec	12	7	07-12-1995	0	773.54	27	3	2	R1013	20.47	5.81	1	0	0

3. **The gender of the patient may be an important factor in determining the cost of hospitalization. The salutations in a beneficiary's name can be used to determine their gender. Make a new field for the beneficiary's gender.**


Formula bar (E2): `=IF(OR(C2="Mr.",C2="Sir",C2="Master"),"Male",IF(OR(C2="Mrs.",C2="Ms.",C2="Miss"),"Female","Unknown"))`

|	A	B	C	D	E	F	G	H	I	J	K	L	M	N	He
1	Custom	Place	Pre-name	Name	Gender	Date	children	charges	Age	Hospital ti	City tier	State ID	BMI	HBA1C	He
2	Id2335	Zywicki	Mr.	Benjamin	Male	09-09-1992	0	563.84	30	2		3 R1013	17.58	4.51	
3	Id2334	Zywek	Mr.	Maciej	Male	30-11-1992	0	570.62	30	2		1 R1013	17.6	4.39	
4	Id2333	Zyryanov	Mr.	Sergey	Male	30-06-1993	0	600	29	2		1 R1013	16.47	6.35	
5	Id2332	Zoeller	Ms.	Lara	Female	13-09-1992	0	604.54	30	3		3 R1013	17.7	6.28	
6	Id2331	Zirdok	Ms.	Lynda	Female	27-07-1998	0	637.26	24	3		3 R1013	22.34	5.57	
7	Id2330	Zimny-Sch	Mr.	Danny	Male	20-11-2001	0	646.14	21	3		3 R1012	22.24	4.29	
8	Id2329	Zimmer	Ms.	Lissa	Female	01-06-1993	0	650	29	3		3 R1013	17.07	5.22	
9	Id2328	Ziegler	Ms.	Sarah	Female	04-07-1995	0	650	27	3		3 R1013	17.82	5.26	
10	Id2327	Zidek	Ms.	Marie	Female	29-11-2002	0	668	20	3		2 R1012	21.77	10.67	
11	Id2326	Zicko	Ms.	Tatyana	Female	09-11-1997	0	670	25	3		3 R1013	20.1	5.6	
12	Id2325	Zhou	Mr.	Ning	Male	12-09-2001	0	687.54	21	3		2 R1013	24.76	4.54	
13	Id2324	Zhou	Mr.	Fan	Male	26-12-1999	0	700	23	3		3 R1013	22.24	5.04	
14	Id2323	Zhang	Mr.	Yu	Male	14-12-1999	0	722.99	23	3		1 R1013	23.35	5.94	
15	Id2322	Zelwin	Mr.	Michael	Male	19-01-2002	0	750	21	3		1 R1012	21.38	8.01	
16	Id2321	Zappala	Mr.	Dante	Male	09-08-1993	0	760	29	3		1 R1013	17.86	5.43	
17	Id2320	Zamore	Mr.	Turner	Male	22-10-1996	0	760	26	3		3 R1013	19.21	5.53	
18	Id2319	Zaferos	Ms.	Audrey	Female	28-06-1993	0	770	29	3		3 R1013	17.08	5.73	
19	Id2318	Zablocki	Mr.	Christoph	Male	18-01-1996	0	770.38	27	3		2 R1012	18.82	5.51	
20	Id2317	Ytterstad	Ms.	Nina	Female	07-12-1995	0	773.54	27	3		2 R1013	20.47	5.81	

4. **State how the distribution is different across gender and tiers of hospitals.**
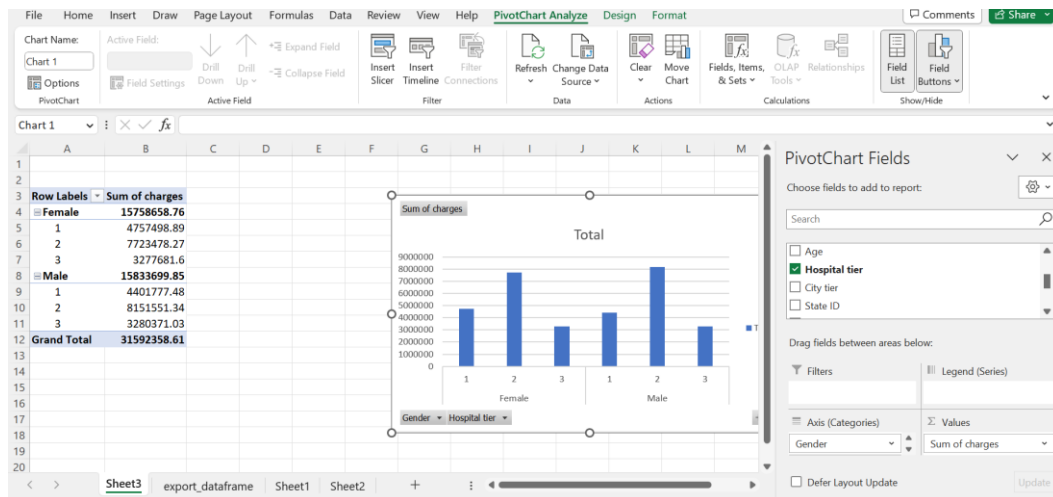
5. Create a radar chart to showcase the median hospitalization cost for each tier of hospitals.
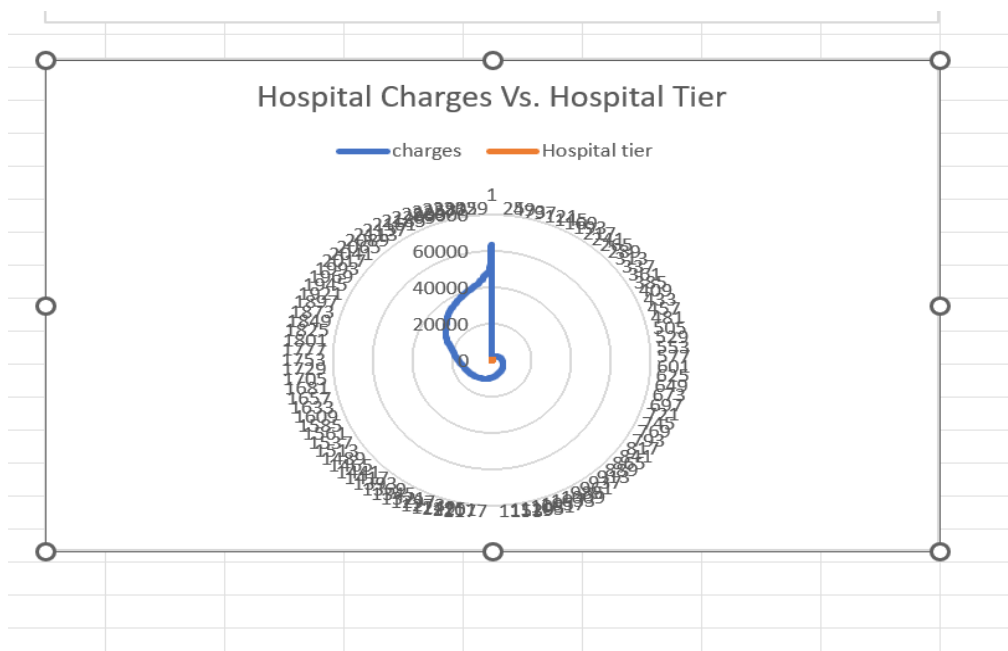
6. Create a frequency table and a stacked bar chart to visualize the count of people in the different tiers of cities and hospitals.
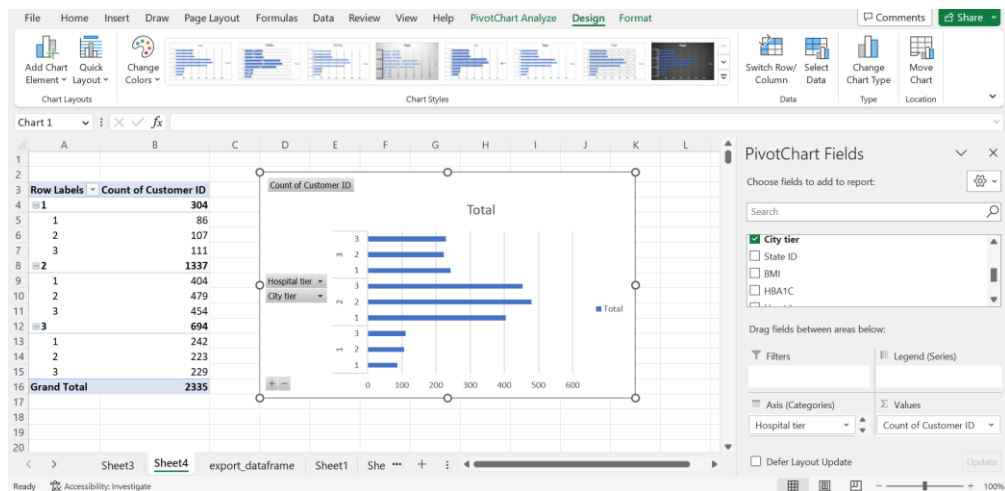
7. Test the following null hypotheses:
  a. The average hospitalization costs for the three types of hospitals are not significantly different

	A	B	C	D	E	F	G	H	I
1	Anova: Single Factor								
2									
3	SUMMARY								
4	*Groups*	*Count*	*Sum*	*Average*	*Variance*				
5	Column 1	2335	5060	2.167024	0.399683				
6	Column 2	2335	31592359	13529.92	1.42E+08				
7									
8									
9	ANOVA								
10	*ce of Varia*	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*		
11	Between (	2.14E+11	1	2.14E+11	3018.161	0	3.843452		
12	Within Gro	3.3E+11	4668	70788987					
13									
14	Total	5.44E+11	4669						
15									
16									
17									
18									

b. The average hospitalization costs for the three types of cities are not significantly different

	A	B	C	D	E	F	G	H	I
1	Anova: Single Factor								
2									
3	SUMMARY								
4	*Groups*	*Count*	*Sum*	*Average*	*Variance*				
5	Column 1	2335	5060	2.167024	0.399683				
6	Column 2	2335	31592359	13529.92	1.42E+08				
7									
8									
9	ANOVA								
10	*ce of Varia*	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*		
11	Between (	2.14E+11	1	2.14E+11	3018.161	0	3.843452		
12	Within Gr	3.3E+11	4668	70788987					
13									
14	Total	5.44E+11	4669						
15									
16									
17									
18									

c. The average hospitalization cost for smokers is not significantly different from the average cost for nonsmokers

F3

	A	B	C	D	E
1	t-Test: Two-Sample Assuming Unequal Variances				
2					
3		*Variable 1*	*Variable 2*		
4	Mean	17019.43	0.06176		
5	Variance	20959480	0.057976		
6	Observati	334	1943		
7	Hypothesi	0			
8	df	333			
9	t Stat	67.94012			
10	P(T<=t) on	1.6E-197			
11	t Critical o	1.649442			
12	P(T<=t) tw	3.2E-197			
13	t Critical tv	1.967113			
14					
15					
16					