In [1]:
```
import pandas as pd
import seaborn as sns; sns.set()
import matplotlib.pyplot as plt
```

In [2]:
```
data = {'Email Type':['Email 1', 'Email 2', 'All Emails (including email
s 1 & 2)'], 'Sends':[2000, 1000, 10000], 'Opens':[1600, 450, 5000]}
```

In [3]:
```
newdf = pd.DataFrame(data)
newdf
```

Out[3]:

	Email Type	Sends	Opens
0	Email 1	2000	1600
1	Email 2	1000	450
2	All Emails (including emails 1 & 2)	10000	5000

In [4]:
```
totalemails = 10000
OtherEmail_Sends = totalemails-2000-1000
OtherEmail_Opens = 5000-450-1600
```

In [5]:
```
otheremails = pd.DataFrame([['Other Emails (Not Email 1 or 2)', OtherEma
il_Sends, OtherEmail_Opens]], columns=['Email Type','Sends','Opens'])
```

In [6]:
```
newdf = pd.concat([newdf, otheremails], ignore_index = True)
newdf
```

Out[6]:

	Email Type	Sends	Opens
0	Email 1	2000	1600
1	Email 2	1000	450
2	All Emails (including emails 1 & 2)	10000	5000
3	Other Emails (Not Email 1 or 2)	7000	2950

In [7]:
```
allrow, otherrow = newdf.iloc[2], newdf.iloc[3]
temp = newdf.iloc[2].copy()
newdf.iloc[2] = otherrow
newdf.iloc[3] = temp
newdf
```

Out[7]:

	Email Type	Sends	Opens
0	Email 1	2000	1600
1	Email 2	1000	450
2	Other Emails (Not Email 1 or 2)	7000	2950
3	All Emails (including emails 1 & 2)	10000	5000

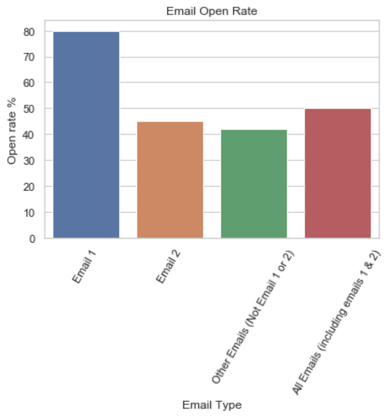
In [8]:
```
newdf['Open rate %'] = newdf['Opens']/newdf['Sends']*100
newdf['# of emails'] = newdf['Sends']/totalemails
newdf
```

Out[8]:

	Email Type	Sends	Opens	Open rate %	# of emails
0	Email 1	2000	1600	80.000000	0.2
1	Email 2	1000	450	45.000000	0.1
2	Other Emails (Not Email 1 or 2)	7000	2950	42.142857	0.7
3	All Emails (including emails 1 & 2)	10000	5000	50.000000	1.0

In [9]:
```
import seaborn as sns
sns.set(style="whitegrid")
ax = sns.barplot(x="Email Type", y="Open rate %", data=newdf)
ax.set_xticklabels(ax.get_xticklabels(), rotation=60)
ax.set_title("Email Open Rate")
```

Out[9]:
```
Text(0.5, 1.0, 'Email Open Rate')
```

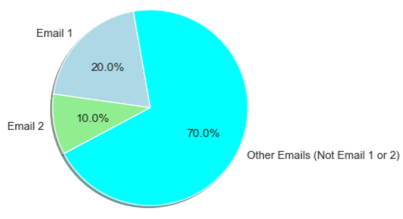


In [10]:
```
import matplotlib.pyplot as plt

# Data to plot
labels = newdf[0:3]['Email Type']
sizes = newdf[0:3]['Sends']
colors = ['lightblue', 'lightgreen', 'cyan']

# Plot
plt.pie(sizes, labels=labels, colors=colors,
autopct='%1.1f%%', shadow=True, startangle=100)

plt.axis('equal')
plt.show()
```



Using OpenRate = Opens/Sends, what insight can you get for the open rates of emails 1 and 2 compared to other emails? Show this on a chart.

The open rate for email 2 (45%) is fairly similar to the open rate across all emails (50%) However the open rate for Email 1 is much more successful, 80% of people open Email 1 type. So we can conclude that Email 1 is almost twice as effective of the recipient opening email 2.

We also have the confidence in this since the sample size for Email 1 is appropriately large. This is shown in the pie chart with the breakdown of the Email types with Email 1 being 20%, and Email 2 being 10% of all the emails sent.

If the sample size of email 1 was too small, we couldn't confidently rely on the ratio as it would reduce the statistical power and increase the margin of error.

What is the Open Rate for all 4 emails combined?