

# 2020 모바일 어플리케이션 마켓 정보 분석

평소에 많이 사용하는 플레이스토어에 배포된 앱을 분석해봤습니다.

## Overview

1. 카테고리별 리뷰 평균 점수
2. 유료앱 여부와 리뷰 점수 상관관계
3. 광고를 포함하는 앱과 그렇지 않은 앱의 리뷰 평균 점수 비교
4. 결론

In [2]:

```
# 필요한 라이브러리 한번에 임포트
import pandas as pd
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
# plot 스타일을 미리 정의해놓은 패키지
import seaborn as sns
import squarify

# 한글 폰트
plt.rcParams['font.family'] = 'AppleGothic'
```

## 공통으로 사용할 플레이 스토어 데이터 읽어오기

In [96]:

```
# 필요한 데이터 임포트
playtore_df = pd.read_csv('playstore.csv', index_col=0, engine='python', encoding='utf-8')
playtore_df.head()
```

Out[96]:

	App Id	Category	Rating	Rating Count	Installs	Minim Inst
App Name						
HTTrack Website Copier	com.httrack.android	Communication	3.6	2848.0	100,000+	10000000
World War 2: Offline Strategy	com.skizze.wwii	Strategy	4.3	17297.0	1,000,000+	10000000
WPSApp	com.themausoft.wpsapp	Tools	4.2	488639.0	50,000,000+	50000000
OfficeSuite - Office, PDF, Word, Excel, PowerPoint	com.mobisystems.office	Business	4.2	1224420.0	100,000,000+	100000000

	App Id	Category	Rating	Rating Count	Installs	Minim Inst
App Name						
<b>Loud Player Free</b>	com.arthelion.loudplayer	Music & Audio	4.2	665.0	50,000+	5000

5 rows × 22 columns

# 1. 카테고리별 리뷰 평균 점수

## 1-1 카테고리별 데이터

In [97]:

```
categories = playtore_df['Category'].unique()
groupedAppByCategory = playtore_df.groupby('Category')

# print('스토어에 등록된 카테고리 목록')
# for category in categories:
#     print(category)
groupedAppByCategory.mean()
```

Out[97]:

	Rating	Rating Count	Minimum Installs	Maximum Installs	Free	Price
Category						
<b>Action</b>	3.199303	40777.040552	1.120423e+06	2.042684e+06	0.957317	0.131585
<b>Adventure</b>	3.115952	11175.093875	3.972175e+05	7.396449e+05	0.918609	0.312795
<b>Arcade</b>	2.766632	14551.488202	6.722659e+05	1.195908e+06	0.957607	0.355522
<b>Art &amp; Design</b>	1.932258	1010.627384	7.422422e+04	1.159923e+05	0.986528	0.071751
<b>Auto &amp; Vehicles</b>	2.275074	1238.954223	2.176497e+05	2.864939e+05	0.961729	0.255783
<b>Beauty</b>	1.700353	503.947622	5.164636e+04	8.358423e+04	0.996151	0.010144
<b>Board</b>	3.152639	9766.717904	4.089541e+05	7.967233e+05	0.936134	0.223015
<b>Books &amp; Reference</b>	2.598829	859.044482	5.295924e+04	1.142776e+05	0.956894	0.260236
<b>Business</b>	1.699900	1095.860691	1.072728e+05	1.542917e+05	0.983985	0.188813
<b>Card</b>	3.420984	9595.768135	3.401958e+05	5.793997e+05	0.940736	0.204947
<b>Casino</b>	3.524858	17468.262239	4.216035e+05	7.367323e+05	0.976208	0.048335
<b>Casual</b>	2.912237	16900.117612	7.048159e+05	1.247210e+06	0.969184	0.076530
<b>Comics</b>	2.988451	2846.556084	1.616816e+05	2.761183e+05	0.974731	0.067272
<b>Communication</b>	2.481514	24513.359748	1.795877e+06	3.321489e+06	0.966016	0.134068
<b>Dating</b>	2.736304	2776.078315	1.413485e+05	2.536160e+05	0.990635	0.172470
<b>Education</b>	2.464080	736.126444	4.106738e+04	7.245474e+04	0.940039	0.328339
<b>Educational</b>	2.781618	2101.643870	3.197827e+05	5.628525e+05	0.902828	0.341438
<b>Entertainment</b>	2.445318	2052.055558	1.454169e+05	2.808652e+05	0.982517	0.069230
<b>Events</b>	1.285573	151.472000	1.364955e+04	2.422356e+04	0.998152	0.024269

	Rating	Rating Count	Minimum Installs	Maximum Installs	Free	Price
Category						
Finance	2.696186	5262.483443	1.626049e+05	2.987099e+05	0.977501	0.132217
Food & Drink	1.483680	1493.822039	5.320995e+04	8.764647e+04	0.989814	0.197240
Health & Fitness	2.087374	2144.082670	1.230789e+05	2.097073e+05	0.960009	0.216481
House & Home	1.548631	1020.150340	6.651043e+04	1.221400e+05	0.990320	0.039560
Libraries & Demo	2.509314	1406.244285	8.681132e+04	1.847807e+05	0.982368	0.125592
Lifestyle	2.368817	1386.768527	8.973138e+04	1.475192e+05	0.975488	0.121770
Maps & Navigation	2.542248	4998.911847	2.122220e+05	3.609678e+05	0.932079	0.477726
Medical	2.182451	689.160571	3.517611e+04	5.851764e+04	0.886650	1.481894
Music	2.674689	8884.830726	5.326958e+05	1.046404e+06	0.952286	0.100370
Music & Audio	2.286137	1620.166055	1.466613e+05	2.296581e+05	0.983222	0.068604
News & Magazines	2.630769	1490.175041	1.943779e+05	3.570050e+05	0.993876	0.033664
Parenting	2.859966	2904.402597	1.027062e+05	2.026065e+05	0.956839	0.125314
Personalization	2.783018	1617.526587	9.686648e+04	1.794558e+05	0.900392	0.170035
Photography	2.477637	7029.124407	4.637618e+05	9.481070e+05	0.977783	0.110174
Productivity	2.327128	4642.518098	8.654924e+05	1.307516e+06	0.940493	0.373952
Puzzle	2.960746	7632.453453	3.516856e+05	6.252575e+05	0.949547	0.176592
Racing	3.250261	27973.368865	1.351167e+06	2.443071e+06	0.968688	0.075548
Role Playing	3.685790	21843.473973	6.654048e+05	1.195284e+06	0.900391	0.452079
Shopping	2.037672	7232.594193	2.431625e+05	4.429796e+05	0.995205	0.041178
Simulation	3.378424	13228.424542	6.071435e+05	1.054192e+06	0.962204	0.125369
Social	2.587373	24510.203117	9.088619e+05	1.607072e+06	0.983988	0.043877
Sports	2.667869	8486.332923	3.091335e+05	5.520891e+05	0.944555	0.448412
Strategy	3.355749	55605.501261	8.956284e+05	1.661757e+06	0.875952	0.414247
Tools	2.690402	6121.086462	8.711995e+05	1.416222e+06	0.939930	0.277288
Travel & Local	2.194510	2496.330514	3.552084e+05	6.271388e+05	0.945131	0.218424
Trivia	2.757646	5797.387186	1.795900e+05	3.172747e+05	0.975320	0.086558
Video Players & Editors	2.901925	22909.516375	1.801543e+06	2.808042e+06	0.958329	0.259368
Weather	3.398638	10823.855635	4.507097e+05	8.682383e+05	0.921907	0.287302
Word	3.227247	10968.786882	3.518533e+05	6.069674e+05	0.959589	0.089838

## 1-2 카테고리 별 앱 개수

In [98]:

```
# avgGroupedAppByCategory = groupedAppByCategory.mean()
```

```

sizegGroupedAppCategory = groupedAppByCategory.size()

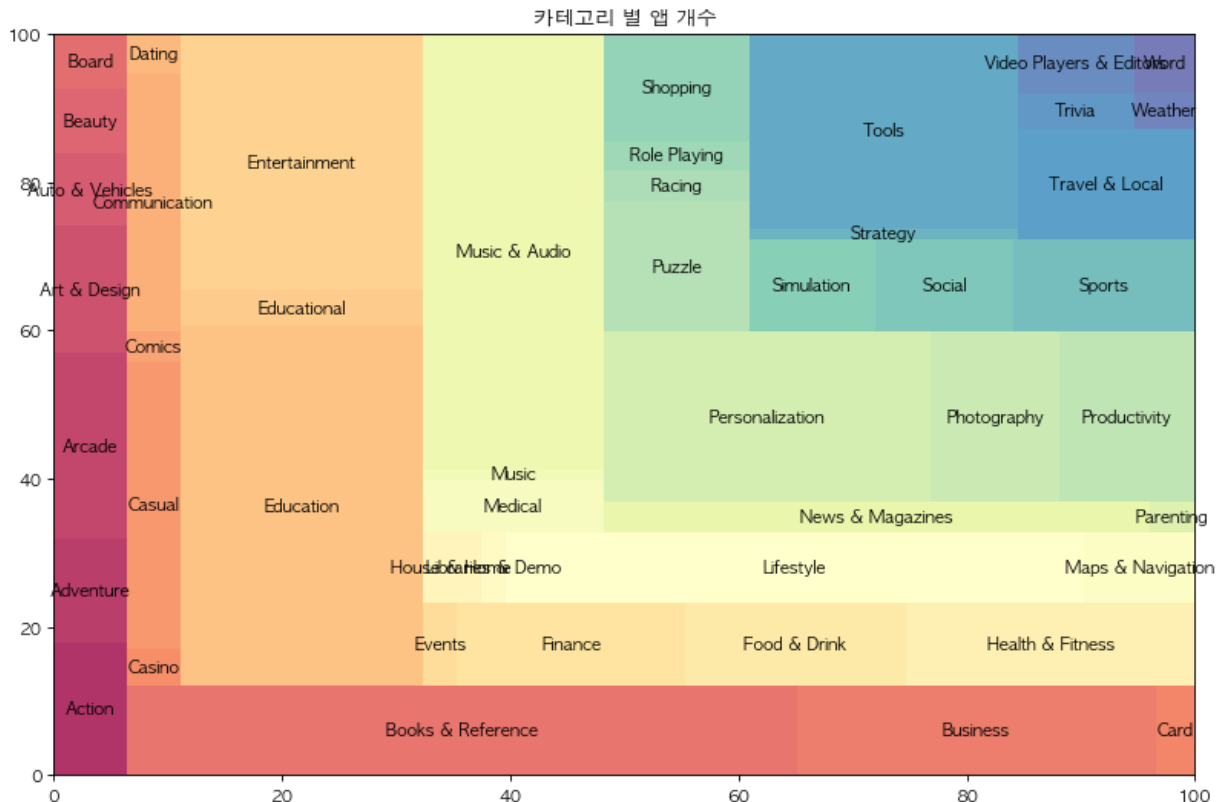
labels = sizegGroupedAppCategory.index.values.tolist()
sizees = sizegGroupedAppCategory
colors = [plt.cm.Spectral(i/float(len(labels))) for i in range(len(labels))]

# Draw Plot
plt.figure(figsize=(12,8), dpi= 80)
squarify.plot(sizes=sizees, label=labels, color=colors, alpha=.8)

# Decorate
plt.title('카테고리 별 앱 개수')

```

Out[98]: Text(0.5, 1.0, '카테고리 별 앱 개수')



In [101]...

```

totalCounts = gizegGroupedAppCategory
totalCounts = gizegGroupedAppCategory.sort_values(axis=0, ascending=False)

upperCategories = totalCounts.iloc[0: 6]
lowerCategories = totalCounts.iloc[-6: ]

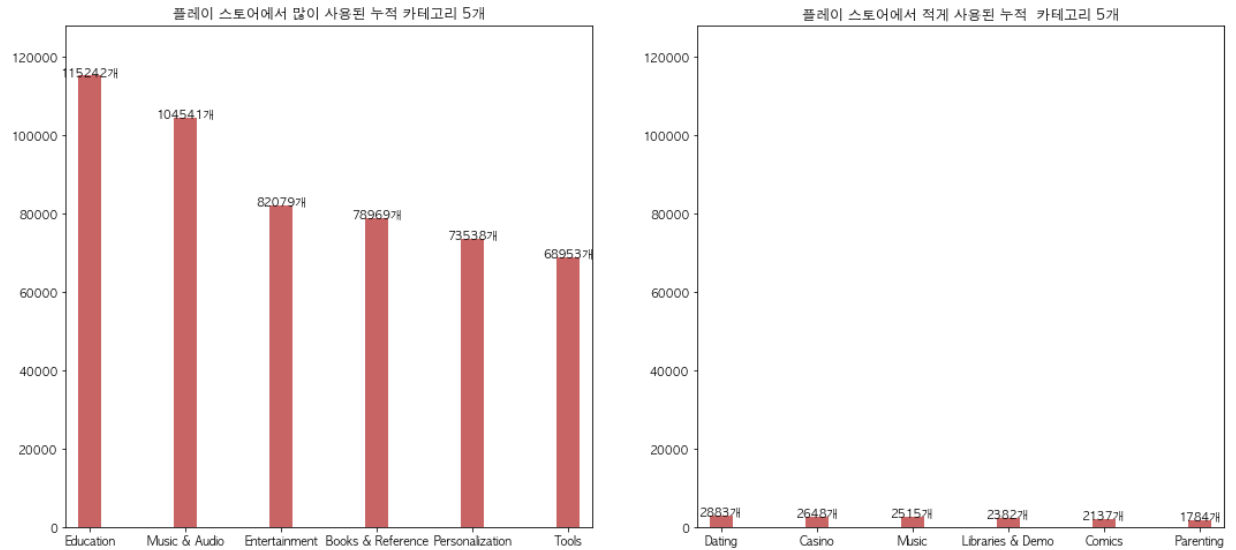
f,ax=plt.subplots(1,2,figsize=(18,8))

ax[0].set_title('플레이 스토어에서 많이 사용된 누적 카테고리 5개')
ax[0].vlines(x=upperCategories.index, ymin=0, ymax=upperCategories.values, color=upperCategories.values)
ax[0].set_ylim([0, upperCategories.values.max()+ upperCategories.values.max()])
for i, value in enumerate(upperCategories.values):
    ax[0].text(i, value+0.5, str(round(value, 1)) + "개", horizontalalignment='center')

ax[1].set_title('플레이 스토어에서 적게 사용된 누적 카테고리 5개')
ax[1].vlines(x=lowerCategories.index, ymin=0, ymax=lowerCategories.values, color=lowerCategories.values)
ax[1].set_ylim([0, highIncidence.values.max() + highIncidence.values.max() / 2])
for i, value in enumerate(lowerCategories.values):
    ax[1].text(i, value+0.5, str(round(value, 1)) + "개", horizontalalignment='center')

plt.show()

```



## 분석 결과

- Education, Music, Entertainment, Books, Personalization, Tools 카테고리는 많은 앱이 배포되어 있음.
- Dating, Casino, Music, Libraries & Demo, Comics, Parenting 카테고리는 적은 앱이 배포되어 있음.
- Education 앱이 의외로 많음
- Dating 앱이 의외로 굉장히 적음

## 1-3 카테고리 별 리뷰 평균 점수 순위

In [103...

```
# Prepare Data
avgGroupCetegory = groupedAppByCategory.mean()
x = avgGroupCetegory['Rating']
x = x.sort_values(axis=0, ascending=True)

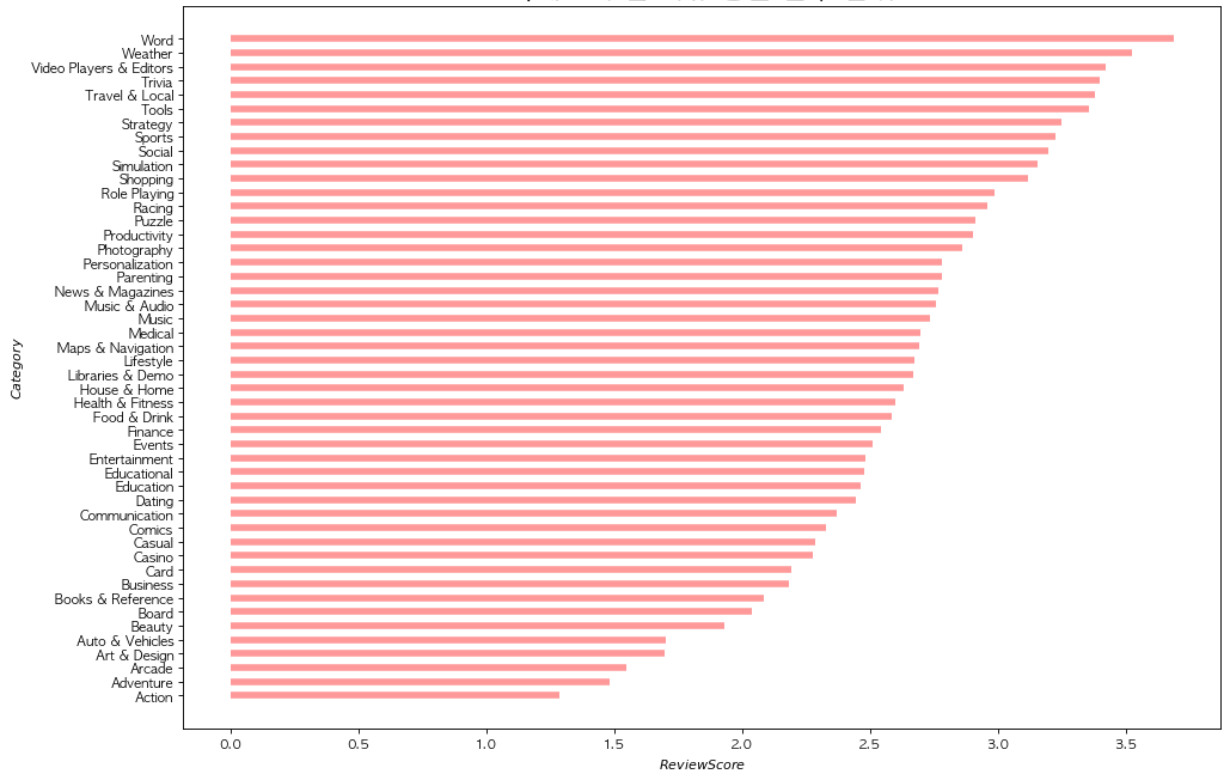
# df['colors'] = ['red' if x < 0 else 'green' for x in df['mpg_z']]
# df.sort_values('mpg_z', inplace=True)
# df.reset_index(inplace=True)

# Draw plot
plt.figure(figsize=(14,10), dpi=80)
plt.hlines(y=avgGroupCetegory.index, xmin=0, xmax=x, color='red', alpha=0.4,

# Decorations
plt.gca().set(ylabel='$Category$', xlabel='$Review Score$')
# plt.yticks(avgGroupCetegory.index, avgGroupCetegory.values, fontsize=12)
plt.title('카테고리 별 리뷰 평균 점수 순위', fontdict={'size':20})
# plt.grid(linestyle='--', alpha=0.5)
```

Out[103... Text(0.5, 1.0, '카테고리 별 리뷰 평균 점수 순위')

카테고리 별 리뷰 평균 점수 순위



## 분석 결과

- Word, Weather, Travel, Social 카테고리의 리뷰 평균 점수가 높음
- Action, Adventure, Arcade, Art & Design, Beauty, Education 카테고리의 평균 점수는 형편없음

## 1-4 최종 분석 결과

- 상위 5 개 앱은 가장 보편적이고, 심사 과정이 까다롭지 않을 거라고 추측할 수 있음, 그 반대는 기피해야 할 카테고리로 추측할 수 있음
- Education 은 배포된 앱은 많은데 리뷰 평균 점수가 형편없음

## 2. 유료앱 여부와 리뷰 점수 상관관계

### 전처리

```
In [57]: groupedAppByFree = playtore_df.groupby('Free')
sizeGroupedAppByFree = groupedAppByFree.size()
```

### 2-1 전체 앱 중 무료앱의 비중

```
In [64]: # Draw Plot
fig, ax = plt.subplots(figsize=(12, 7), subplot_kw=dict(aspect="equal"), dpi=
explode = [0,0]
def func(pct, allvals):
    absolute = int(pct/100.*np.sum(allvals))
    return "{:.1f}% ({:d})".format(pct, absolute)

wedges, texts, autotexts = ax.pie(sizeGroupedAppByFree,
                                autopct=lambda pct: func(pct, sizeGroupedAppByFree))
```

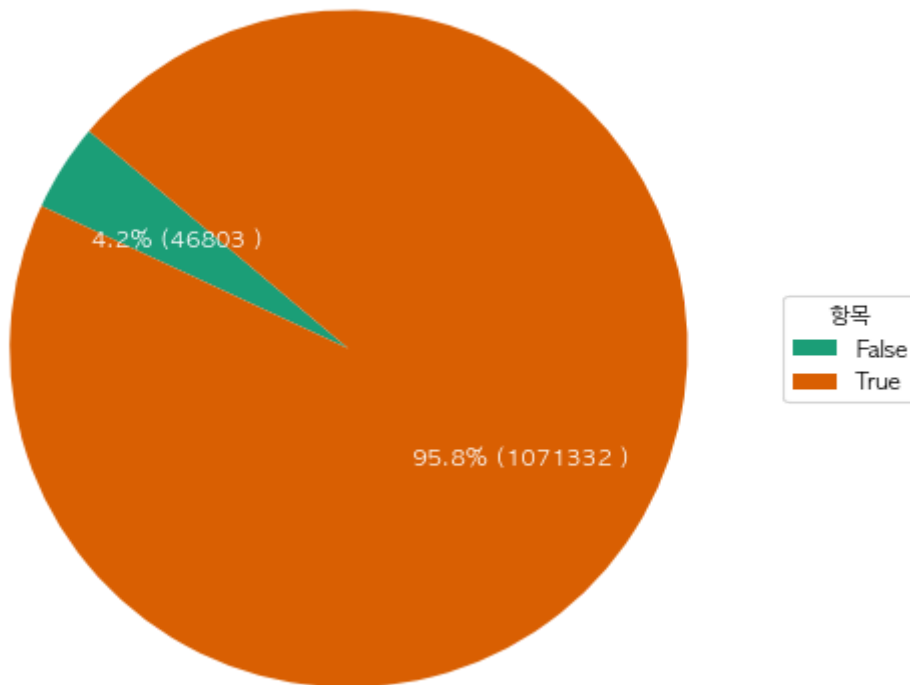
```

textprops=dict(color="w"),
colors=plt.cm.Dark2.colors,
startangle=140,
explode=explode)

# Decoration
ax.legend(wedges, sizeGroupedAppByFree.index, title="항목", loc="center left",
plt.setp(ax.legend().texts, size=10, weight=700)
ax.set_title("전체 앱 중 무료앱 비중")
plt.show()

```

전체 앱 중 무료앱 비중



## 2-2 유료앱과 무료앱 평균 리뷰 점수

```
In [66]: groupedAppByFree.mean()['Rating']
```

```
Out[66]: Free
False    2.148196
True     2.505278
Name: Rating, dtype: float64
```

## 2-3 최종 분석 결과

- 무료앱이 전체 앱 중 95.6% 를 차지함.
- 무료 앱(2.50)이 유료 앱(2.14)보다 평균 리뷰 점수가 높음.

## 3. 광고를 포함하는 앱과 그렇지 않은 앱의 리뷰 평균 점수 비교

```
In [75]: groupedAppByAd = playtore_df.groupby('Ad Supported')
```

```
sizeGroupedAppByAd = groupedAppByAd.size()
sizeGroupedAppByAd
```

Out[75]: <pandas.core.groupby.generic.DataFrameGroupBy object at 0x18f54a550>

### 3-1 전체 앱 중 광고를 포함하는 앱의 비중

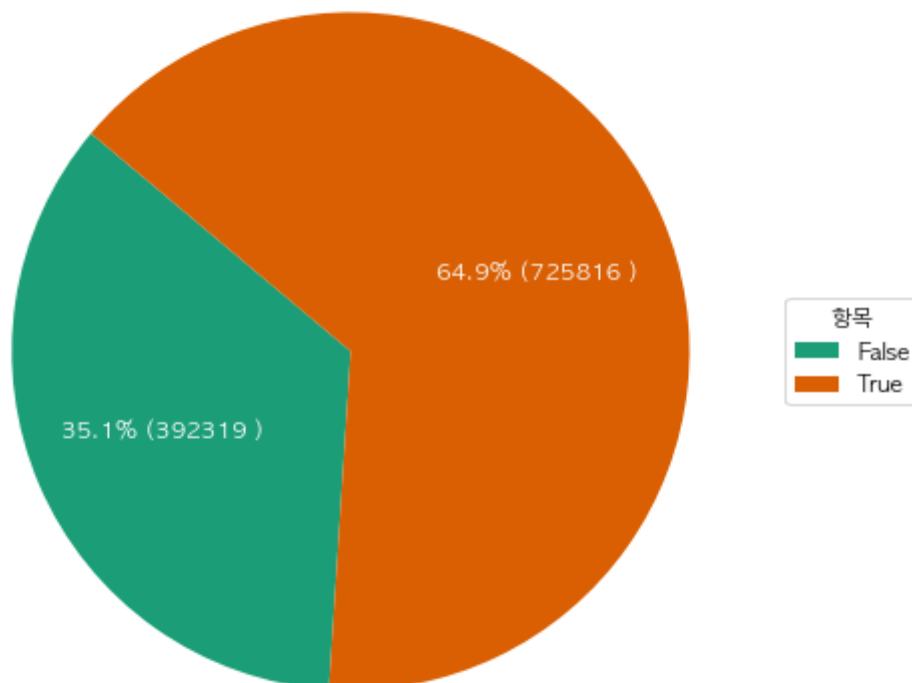
In [104...

```
# Draw Plot
fig, ax = plt.subplots(figsize=(12, 7), subplot_kw=dict(aspect="equal"), dpi=
explode = [0,0]
def func(pct, allvals):
    absolute = int(pct/100.*np.sum(allvals))
    return "{:.1f}% ({:d} )".format(pct, absolute)

wedges, texts, autotexts = ax.pie(sizeGroupedAppByAd,
                                autopct=lambda pct: func(pct, sizeGroupedAppByAd.index),
                                textprops=dict(color="w"),
                                colors=plt.cm.Dark2.colors,
                                startangle=140,
                                explode=explode)

# Decoration
ax.legend(wedges, sizeGroupedAppByAd.index, title="항목", loc="center left", b
plt.setp(autotexts, size=10, weight=700)
ax.set_title("전체 앱 중 광고를 포함하는 앱의 비중")
plt.show()
```

전체 앱 중 광고를 포함하는 앱의 비중



### 3-2 광고를 포함한 앱과 그렇지 않은 앱 리뷰 평균 점수 비교

In [80]:

```
groupedAppByAd.mean()['Rating']
```



```
Out[80]: Ad Supported
False    2.221540
True     2.634685
Name: Rating, dtype: float64
```

### 3-3 최종 분석 결과

- 광고를 포함하는 앱이 전체의 64%임
- 광고를 포함하는 앱이 오히려 리뷰 평균 점수가 높았다.

### 분석최종 정리

- Education, Music, Entertainment, Books, Personalization, Tools 카테고리는 비교적 배포 심사의 허들이 많을 수 있음.
- Dating, Casino, Music, Libraries & Demo, Comics, Parenting 카테고리는 앱이 심사에서 거절 당할 확률이 높을 있음.
- 유료앱보다는 무료앱의 리뷰 평균 점수가 높음
- 광고 포함 여부는 생각보다 리뷰 점수에 크게 영향을 끼치지 않음.