

회원 수 분석 및 상품 종류별 분포

1. 프로젝트 개요

- **프로젝트 제목:** 회원 수 분석 및 상품 종류별 분포
- **프로젝트 목적:** 회원 수 데이터를 분석하여 유료 및 무료 상품의 분포를 이해하고, 이를 바탕으로 향후 전략 수립에 도움을 주고자 함

2. 데이터 수집 및 준비

- **데이터 출처:** 회사 내부 데이터베이스 (샘플 데이터 사용)
- **데이터 파일 형식:** Excel
- **데이터 준비:** 데이터를 불러오고 필요한 전처리를 수행

3. 데이터 분석 과정

데이터 불러오기 및 전처리

```
import pandas as pd
```

```
In [259...]: learn = pd.read_excel('1.전체회원수.xlsb', engine='pyxlsb')
```

```
In [226...]: learn
learn.shape
learn.info()
learn.describe()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 116699 entries, 0 to 116698
Data columns (total 52 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   학생명          116699 non-null   object 
 1   학생아이디      116699 non-null   object 
 2   학생SID         116699 non-null   int64  
 3   팀명            115313 non-null   object 
 4   파트명          115700 non-null   object 
 5   관리교사명      115726 non-null   object 
 6   그룹            116699 non-null   object 
 7   학년            116699 non-null   object 
 8   학습계획시작일  103949 non-null   float64
 9   학습계획종료일  103949 non-null   object 
 10  최근계획수정일 103949 non-null   object 
 11  단말기구분      0 non-null      float64
 12  상품명          116677 non-null   object 
 13  현재상품주문번호 116699 non-null   int64  
 14  학습시작일      116699 non-null   int64  
 15  학습마감일      116699 non-null   int64  
 16  학부모집전화      35877 non-null   object 
 17  학부모핸드폰번호 116699 non-null   int64  
 18  학생핸드폰번호  113719 non-null   object 
 19  휴회여부        116699 non-null   object 
 20  학부모주소       116107 non-null   object 
 21  학부모상세주소  115716 non-null   object 
 22  학부모앱사용여부 116699 non-null   object 
 23  시리얼번호      0 non-null      float64
 24  최종맥로그인시리얼 81268 non-null   object 
 25  080수신여부 Y:수신거부자 116699 non-null   object 
 26  우편번호2       116093 non-null   object 
 27  student_zipcode 70427 non-null   object 
 28  student_주소     71433 non-null   object 
 29  학생상세주소    71229 non-null   object 
 30  월결제 상태      114489 non-null   object 
 31  결제 날짜        116677 non-null   float64
 32  유료상품 구매 개수 116699 non-null   int64  
 33  미납횟수        116677 non-null   float64
 34  학부모우편번호  116093 non-null   object 
 35  sms수신여부    116699 non-null   object 
 36  학교명          116699 non-null   object 
 37  교사아이디      115726 non-null   object 
 38  학부모성함      116699 non-null   object 
 39  학부모아이디    116699 non-null   object 
 40  휴지여부        116699 non-null   object 
 41  휴지 종료일     10580 non-null   float64
 42  휴지여부(과거)  116699 non-null   object 
 43  휴지종료일(과거) 0 non-null      float64
 44  체험상품여부N:유료 116677 non-null   object 
 45  최초결제 학습시작일 113443 non-null   float64
 46  주문상태        116677 non-null   object 
 47  학습상태        116699 non-null   object 
 48  상품종류        116699 non-null   object 
 49  해외학습여부    116699 non-null   object 
 50  기기타입        105782 non-null   object 
 51  관리유형        116699 non-null   object 

dtypes: float64(8), int64(6), object(38)
memory usage: 46.3+ MB

```

Out[226...]

	학생SID	학습계획시작일	단말기구분	현재상품주문번호	학습시작일	학습마감일
count	1.166990e+05	103949.000000	0.0	1.166990e+05	116699.000000	116699.000000
mean	2.225547e+06	44927.743999	NaN	3.864194e+16	44947.077876	45493.813477
std	4.005250e+05	103.589109	NaN	2.248865e+16	205.692851	221.272911
min	9.300000e+01	43953.000000	NaN	3.100072e+13	42831.000000	45231.000000
25%	2.024190e+06	44867.000000	NaN	2.307260e+15	44853.000000	45333.000000
50%	2.319470e+06	44867.000000	NaN	5.209260e+16	44988.000000	45428.000000
75%	2.532526e+06	44961.000000	NaN	5.304240e+16	45105.000000	45646.000000
max	2.716371e+06	45232.000000	NaN	5.310310e+16	45231.000000	73055.000000

In [227...]: learn['상품종류'].dtypes

Out[227...]: dtype('O')

In [228...]: learn['상품종류'].value_counts()

Out[228...]: 상품종류
 유료 114035
 무료 2664
 Name: count, dtype: int64

In [229...]: learn['상품종류'].isna().sum()

Out[229...]: 0

유료 및 무료 상품 학습 회원수 구하기

```
In [ ]: import seaborn as sns
import matplotlib.pyplot as plt
import matplotlib.font_manager as fm

font_path = 'malgun.ttf'
fontprop = fm.FontProperties(fname=font_path, size=10)
plt.rc('font', family=fontprop.get_name())

# 유료 및 무료 상품 그래프 그리기
plt.figure(figsize=(8, 6))
sns.countplot(data=learn, x='상품종류')
plt.xlabel('상품종류', fontproperties=fontprop)
plt.ylabel('Count', fontproperties=fontprop)
plt.title('유료 및 무료 상품의 개수', fontproperties=fontprop)
plt.show()
```

데이터 정제

```
In [263...]: # 상품명 중, 교사 테스트 상품 제외
f_learn = learn[~learn['상품명'].str.contains('교사용', na=False)]
```

```
In [233...]: f_learn['상품종류'].value_counts()
```

```
Out[233...]: 상품종류
유료      112855
무료      1691
Name: count, dtype: int64
```

```
In [273...]: # 학생 이름 중, 테스트가 포함된 회원 제외
```

```
f2_learn = f_learn[~f_learn['학생명'].str.contains('테스트|test', case=False, na=False)]
```

```
In [283...]: f2_learn['상품종류'].value_counts()
```

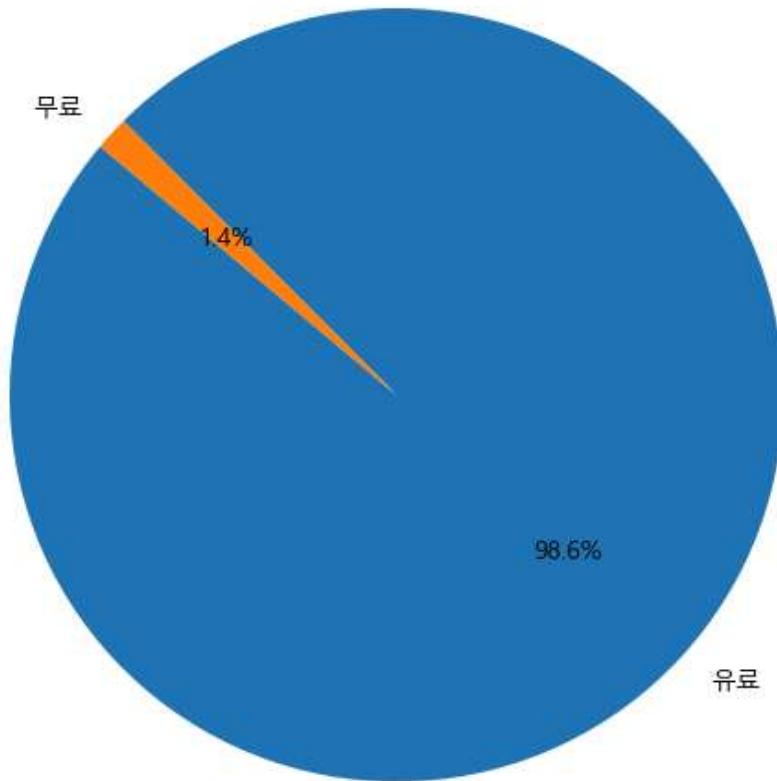
```
Out[283...]: 상품종류
유료      112359
무료      1619
Name: count, dtype: int64
```

데이터 시각화

```
In [285...]: product_counts = f2_learn['상품종류'].value_counts()
```

```
In [277...]: # 교사 테스트 상품 제외, 기타 테스트 학습 회원 제외 후 실제 학습 중인 학생들의 상품 종류
plt.figure(figsize=(8, 6))
plt.pie(product_counts, labels=product_counts.index, autopct='%1.1f%%', startangle=90)
plt.title('유료 및 무료 상품의 비율', fontproperties=fontprop)
plt.axis('equal')
plt.show()
```

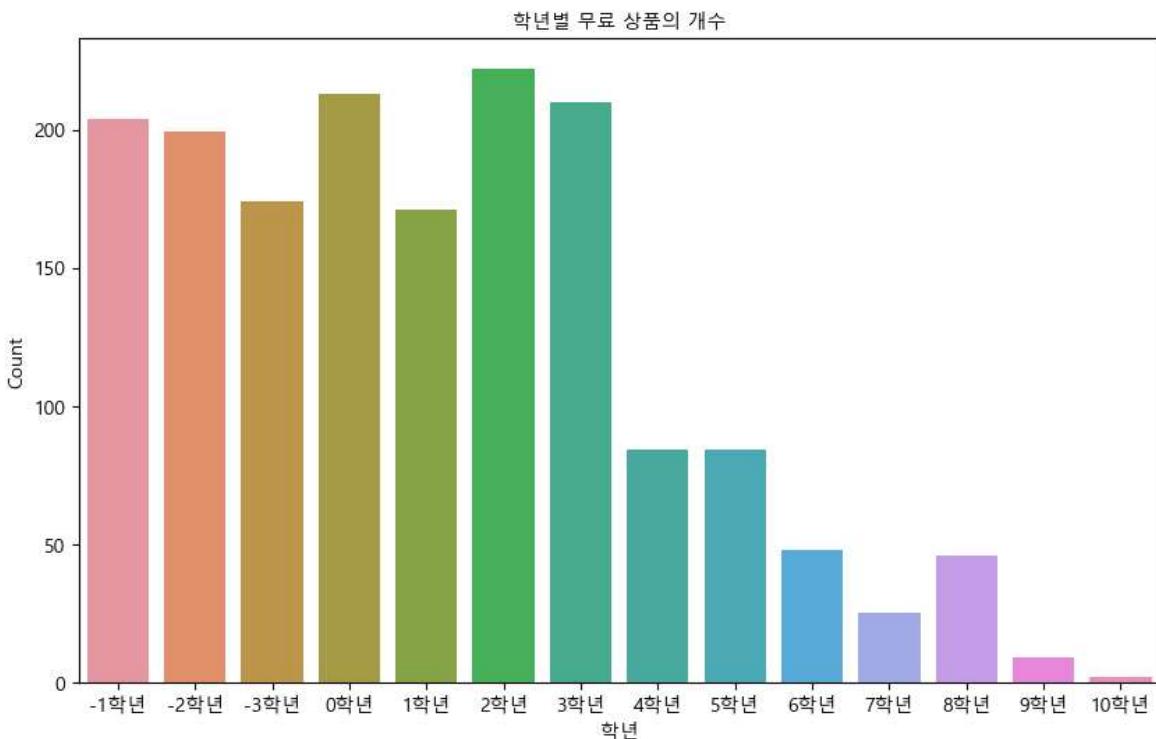
유료 및 무료 상품의 비율



무료 상품 학습 데이터 시각화(학년별, 팀별)

```
In [287...]: # 무료 상품 학습중인 회원들만 필터링  
experience = f2_learn[f2_learn['상품종류'] == '무료']  
  
In [237...]: # 학년별 무료 상품 학습 회원 그래프 그리기  
unique_grades = sorted([grade for grade in experience['학년'].unique() if grade in experience['학년']] = pd.Categorical(experience['학년'], categories=unique_grades)  
  
plt.figure(figsize=(10, 6))  
sns.countplot(data=experience, x='학년')  
plt.xlabel('학년', fontproperties=fontprop)  
plt.ylabel('Count', fontproperties=fontprop)  
plt.title('학년별 무료 상품의 개수', fontproperties=fontprop)  
plt.show()
```

```
C:\Users\Public\Documents\ESTsoft\CreatorTemp\ipykernel_22324\772267901.py:2: SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame.  
Try using .loc[row_indexer,col_indexer] = value instead  
  
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy  
experience['학년'] = pd.Categorical(experience['학년'], categories=unique_grades, ordered=True)  
C:\Users\seyou\anaconda3\Lib\site-packages\seaborn\categorical.py:641: FutureWarning: The default of observed=False is deprecated and will be changed to True in a future version of pandas. Pass observed=False to retain current behavior or observed=True to adopt the future default and silence this warning.  
grouped_vals = vals.groupby(grouper)
```



In [289...]: # 팀별 무료 체험 회원수 구하기

```
experience['팀명'].value_counts()
```

```
Out[289...]: 팀명
상담1센터      581
상담2센터      490
홈런방문3센터   44
홈런방문5센터   42
홈런방문6센터   40
홈런방문1센터   32
홈런방문4센터   30
홈런방문2센터   23
홈런방문7센터   19
학습지원실(기타) 19
3지부          15
SL사업실_스탭    11
경남지사        11
1지부          10
광주전남지사     8
영업기획        6
지사영업실       6
상담센터(해피콜) 6
리틀 관리        4
2지부          4
구미김천지사     3
고양파주지사     3
대전지사        3
부산북부지사     3
상담영업관리팀   3
경기서부지사     2
강릉지사(CP)     2
기장양산지사     2
인천남부지사     2
제주지사(CP)     2
인천북부지사     2
안동영주지사     2
김포지사        2
남동지사        2
서울동부지사     2
부산남부지사     2
서울강서지사     2
천안지사        2
중등 관리        1
동부지사        1
지사영업실_스탭 1
[내부]마케팅     1
동북지사        1
방문영업관리팀(직원) 1
HAMS_BETA_TEST팀 1
울산지사        1
대구동북지사     1
초등운영팀      1
Name: count, dtype: int64
```

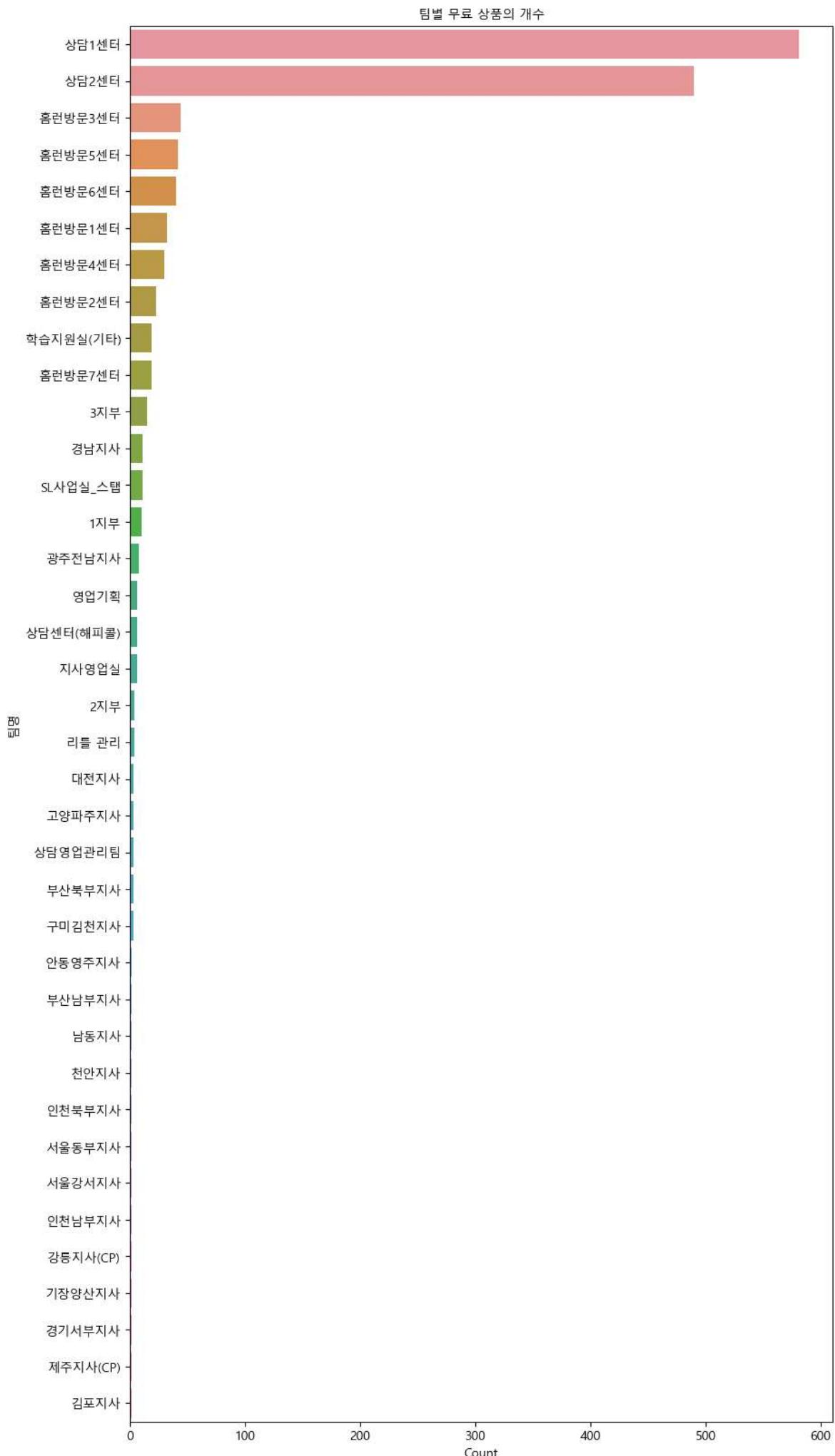
```
filter_team['팀명'].value_counts()
```

```
In [311...]: # '팀명'별 빈도 계산
team_counts = experience['팀명'].value_counts()

# 빈도가 2이상인 팀만 필터링
teams_with_more_than_two = team_counts[team_counts > 1].index
filtered_experience = experience[experience['팀명'].isin(teams_with_more_than_tw]
```

```
# 빈도 데이터 프레임 생성
team_counts_df = filtered_experience['팀명'].value_counts().reset_index()
team_counts_df.columns = ['팀명', 'Count']

# 그래프 그리기
plt.figure(figsize=(10, 20))
sns.barplot(data=team_counts_df, y='팀명', x='Count')
plt.ylabel('팀명', fontproperties=fontprop)
plt.xlabel('Count', fontproperties=fontprop)
plt.title('팀별 무료 상품의 개수', fontproperties=fontprop)
plt.show()
```



4. 결과 해석

결과 요약

1. 1학년부터 3학년까지 무료 상품 학습 회원이 많음.

1학년, 2학년, 3학년의 무료 상품 개수가 다른 학년에 비해 상대적으로 많습니다. 이는 초등 저학년 학생들을 대상으로 한 무료 상품이 많이 제공되고 있음을 나타냅니다.

2. 고학년으로 갈수록 무료 상품 학습 회원 감소

4학년 이후부터 학년이 올라갈수록 무료 상품의 개수가 감소하는 경향이 보입니다. 특히, 6학년 이후부터는 급격히 감소하여 10학년에서는 거의 무료 상품이 제공되지 않습니다.

3. 상담1센터와 상담2센터의 무료 상품 학습 회원수

상담1센터와 상담2센터의 무료 상품 학습 회원수가 다른 센터에 비해 매우 많습니다. 특히 상담1센터는 가장 많은 회원수를 가지고 있으며, 상담2센터가 그 뒤를 잇고 있습니다. 이는 이 두 센터가 활발하게 활동하고 있음을 시사합니다.

4. 나머지 센터들의 무료 상품 학습 회원수

나머지 센터들의 회원은 상대적으로 적습니다. 상담1센터와 상담2센터에 비해 현저히 적은 회원수를 가지고 있습니다. 여러 센터들이 비슷한 수준의 회원을 가지고 있지만, 큰 차이는 보이지 않습니다.

5. 회원수가 적은 센터

데이터 개수가 특히 적은 센터들은 활동이 저조하거나 데이터 수집이 미흡할 수 있습니다.

인사이트

1. 초등 저학년 타겟 무료 상품 집중

무료 상품이 1학년부터 3학년 사이에 집중되어 있다는 것은 이 연령대가 중요한 타겟 그룹임을 시사합니다. 이 연령대의 학생들이 무료 상품으로 많이 학습을 하고, 해당 연령대의 자녀를 둔 학부모들이 무료 상품 학습에 관한 관심이 많은 것을 알 수 있습니다. 이에 맞추어 저학년을 타겟으로 한 무료 상품에 집중해야 함을 알 수 있습니다.

2. 고학년 학생을 위한 무료 상품 부족

고학년 학생들을 위한 무료 상품이 상대적으로 부족하다는 점은 보완이 필요합니다. 특히, 6학년 이상 학생들을 대상으로 한 무료 상품의 개발이나 제공을 고려할 수 있습니다.

3. 상담센터의 중요성

상담1센터와 상담2센터의 회원수가 매우 많은 것으로 보아, 이 센터들은 주요 활동 센터임을 알 수 있습니다. 이 센터들이 중요한 역할을 하고 있으며, 많은 데이터가 이곳에서 발생하고 있음을 나타냅니다.

4. 데이터 수집의 편차

센터별로 데이터 수집에 큰 편차가 있는 것으로 보입니다. 상담센터를 제외한 다른 센터들의 데이터 수집이 상대적으로 적다는 점에서, 데이터 수집 방법이나 기준에 일관성이 없을 가능성이 있습니다.

5. 데이터가 적은 센터들에 대한 분석

데이터가 적은 센터들은 왜 데이터가 적은지에 대한 추가 분석이 필요합니다. 활동이 적은 것인지, 아니면 데이터 수집이 미흡한 것인지 파악해야 합니다.

5. 향후 계획

1. 고학년 대상 무료 상품 개발

6학년 이상 학생들을 위한 무료 상품을 개발하여, 다양한 연령대의 학생들에게 혜택을 제공합니다. 고학년 학생들이 관심을 가질 만한 콘텐츠나 서비스를 조사하고, 이를 바탕으로 상품을 기획합니다.

2. 저학년 타겟 마케팅 강화

1학년부터 3학년까지의 학생들을 대상으로 한 마케팅을 강화합니다. 무료 상품의 효과적인 활용 방안을 제시하고, 학부모들에게 적극적으로 홍보합니다.

3. 데이터 수집 방법 개선

센터별로 데이터 수집 방법을 표준화하여 데이터 수집의 일관성을 높입니다. 특히, 회원수가 적은 센터들을 대상으로 데이터 수집 방법을 개선할 필요가 있습니다.

4. 상담센터의 활동 강화

상담1센터와 상담2센터의 데이터를 활용하여 성공 사례를 다른 센터에 전파합니다. 이러한 사례를 바탕으로 다른 센터의 활동을 강화할 수 있는 방안을 모색합니다.

5. 데이터 분석 강화

각 센터의 데이터 분석을 통해, 어떤 활동이 데이터 증가에 기여하는지 파악합니다. 데이터를 기반으로 한 의사결정을 통해 센터들의 활동을 최적화합니다.

6. 추가 조사 및 분석

데이터가 적은 센터들을 대상으로 추가 조사를 실시하여, 데이터 수집이 적은 이유를 파악합니다. 필요하다면 데이터 수집 인프라를 개선합니다.