University Admissions in Turkey (2019–2024)

Scholarship level, score type, and demand per quota

Problem topic

This notebook explores how **scholarship type** (e.g., Ücretsiz, Burslu, %50 İndirimli, Ücretli) and **score type** (SAY, EA, SÖZ, DİL, TYT) relate to **demand per quota** across Turkish university programs from 2019–2024.

Central question

How has **demand per quota** evolved over time by scholarship level, and do these patterns differ across score types and regions (cities)?

Dataset overview

Each row represents a **university program in a specific year**. The dataset includes:

- program identifiers (`program_code`)
- program context (`university_name` , `department_name` , `faculty_name` , `city` , `university_type`)
- admissions demand & outcomes (`total_preferences` , `demand_per_quota` , `placed_count` , `initial_placement_rate`)
- scoring competitiveness (`final_score_012` , `final_rank_012` , `final_score_018` , `final_rank_018`)
- student composition (`male` , `female`)
- scholarship categories (`scholarship_type`)
- score category (`score_type`)

This is the assignment file for Statistics module Everything Counts at LIS. Student Number 25000148737, Full-Time MASc 2025.

Available on GitHub at: https://github.com/tulin-b/everything-counts-assignment1-25000148737.git

In [198…
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import scipy.stats as stats
```

In [199…
```
print(pd)
```

```
<module 'pandas' from '/opt/anaconda3/lib/python3.13/site-packages/pandas/__init__.py'>
```

In [200…
```
df = pd.read_csv("01_university_admissions_turkey_2019_2024.csv")
df.head()
```

Out[200…

	program_code	year	university_name	city	university_type	department_name	faculty_n
0	106510077	2019	ABDULLAH GÜL ÜNİVERSİTESİ	KAYSERİ	devlet	Bilgisayar Mühendisliği	Mühenc Fakü
1	106510077	2020	ABDULLAH GÜL ÜNİVERSİTESİ	KAYSERİ	devlet	Bilgisayar Mühendisliği	Mühenc Fakü
2	106510077	2021	ABDULLAH GÜL ÜNİVERSİTESİ	KAYSERİ	devlet	Bilgisayar Mühendisliği	Mühenc Fakü
3	106510077	2022	ABDULLAH GÜL ÜNİVERSİTESİ	KAYSERİ	devlet	Bilgisayar Mühendisliği	Mühenc Fakü
4	106510077	2023	ABDULLAH GÜL ÜNİVERSİTESİ	KAYSERİ	devlet	Bilgisayar Mühendisliği	Mühenc Fakü

5 rows × 39 columns

In [201…

```
df.tail()
```

Out[201…

	program_code	year	university_name	city	university_type	department_name	facu
128347	109370154	2022	ŞIRNAK ÜNİVERSİTESİ	ŞIRNAK	devlet	İnşaat Teknolojisi	Şırn Yü
128348	109370154	2023	ŞIRNAK ÜNİVERSİTESİ	ŞIRNAK	devlet	İnşaat Teknolojisi	Şırn Yü
128349	109370154	2024	ŞIRNAK ÜNİVERSİTESİ	ŞIRNAK	devlet	İnşaat Teknolojisi	Şırn Yü
128350	109310025	2019	ŞIRNAK ÜNİVERSİTESİ	ŞIRNAK	devlet	İşletme	İd
128351	109310025	2024	ŞIRNAK ÜNİVERSİTESİ	ŞIRNAK	devlet	İşletme	İd

5 rows × 39 columns

Initial structure check

After loading that data, I checked the size, variable types and more to understand the context better:

- what each row represents
- which columns are numeric vs categorical
- any immediate cleaning needs

In [202…

```
print("Shape:", df.shape)
display(df.info())
display(df.describe(include="all").T)
```

```
Shape: (128352, 39)
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 128352 entries, 0 to 128351
Data columns (total 39 columns):
 #   Column                          Non-Null Count   Dtype
---  ------                          --------------   -----
 0   program_code                    128352 non-null  int64
 1   year                            128352 non-null  int64
 2   university_name                 128352 non-null  object
 3   city                            128352 non-null  object
 4   university_type                 128352 non-null  object
 5   department_name                 128352 non-null  object
 6   faculty_name                    127475 non-null  object
 7   score_type                      128352 non-null  object
 8   scholarship_type                128352 non-null  object
 9   is_undergraduate                128352 non-null  bool
 10  all_tags                        70270 non-null   object
 11  total_quota                     128352 non-null  int64
 12  total_enrolled                  128352 non-null  int64
 13  male                            128352 non-null  int64
 14  female                          128352 non-null  int64
 15  final_score_012                 115482 non-null  float64
 16  final_rank_012                  114116 non-null  float64
 17  final_score_018                 50153 non-null   float64
 18  final_rank_018                  49190 non-null   float64
 19  initial_placement_rate          128352 non-null  float64
 20  not_registered                  128352 non-null  int64
 21  additional_placement            128352 non-null  int64
 22  avg_obp_012                     124227 non-null  float64
 23  avg_obp_018                     67262 non-null   float64
 24  total_preferences               128352 non-null  int64
 25  demand_per_quota                128352 non-null  float64
 26  avg_preference_rank             128352 non-null  float64
 27  top_1_pref_count                128352 non-null  int64
 28  top_3_pref_count                128352 non-null  int64
 29  top_9_pref_count                128352 non-null  int64
 30  placed_count                    128352 non-null  int64
 31  placed_pref_rank_avg            124363 non-null  float64
 32  placed_top_1_pref_count         128352 non-null  int64
 33  placed_top_3_pref_count         128352 non-null  int64
 34  placed_top_10_pref_count        128352 non-null  int64
 35  placed_pref_uni_devlet_count    128352 non-null  int64
 36  placed_pref_uni_vakif_count     128352 non-null  int64
 37  placed_pref_uni_kktc_count      128352 non-null  int64
 38  placed_pref_uni_yurt_disi_count  128352 non-null  int64
dtypes: bool(1), float64(10), int64(20), object(8)
memory usage: 37.3+ MB
None
```

	count	unique	top	freq	mean	
program_code	128352.0	NaN	NaN	NaN	154556522.746229	61566
year	128352.0	NaN	NaN	NaN	2021.500569	
university_name	128352	235	İSTANBUL GELİŞİM ÜNİVERSİTESİ	2251	NaN	
city	128352	82	İSTANBUL	39416	NaN	
university_type	128352	4	devlet	73074	NaN	
department_name	128352	733	Bilgisayar Programcılığı	2388	NaN	
faculty_name	127475	1131	Sağlık Hizmetleri Meslek Yüksekokulu	10825	NaN	
score_type	128352	5	TYT	60030	NaN	
scholarship_type	128352	10	Ücretsiz	63799	NaN	
is_undergraduate	128352	2	True	68322	NaN	
all_tags	70270	83	Burslu	14962	NaN	
total_quota	128352.0	NaN	NaN	NaN	47.489786	
total_enrolled	128352.0	NaN	NaN	NaN	44.557233	
male	128352.0	NaN	NaN	NaN	20.590119	
female	128352.0	NaN	NaN	NaN	23.796707	
final_score_012	115482.0	NaN	NaN	NaN	291.356824	
final_rank_012	114116.0	NaN	NaN	NaN	756699.852019	6
final_score_018	50153.0	NaN	NaN	NaN	266.989731	
final_rank_018	49190.0	NaN	NaN	NaN	1075539.606444	539
initial_placement_rate	128352.0	NaN	NaN	NaN	89.225721	
not_registered	128352.0	NaN	NaN	NaN	4.741928	
additional_placement	128352.0	NaN	NaN	NaN	3.614801	
avg_obp_012	124227.0	NaN	NaN	NaN	384.429394	
avg_obp_018	67262.0	NaN	NaN	NaN	281.277206	
total_preferences	128352.0	NaN	NaN	NaN	877.70203	3
demand_per_quota	128352.0	NaN	NaN	NaN	18.744385	
avg_preference_rank	128352.0	NaN	NaN	NaN	9.256056	
top_1_pref_count	128352.0	NaN	NaN	NaN	69.302894	
top_3_pref_count	128352.0	NaN	NaN	NaN	200.026443	1
top_9_pref_count	128352.0	NaN	NaN	NaN	512.74859	2
placed_count	128352.0	NaN	NaN	NaN	44.386827	
placed_pref_rank_avg	124363.0	NaN	NaN	NaN	5.697212	
placed_top_1_pref_count	128352.0	NaN	NaN	NaN	11.317572	
placed_top_3_pref_count	128352.0	NaN	NaN	NaN	22.155338	

	count	unique	top	freq	mean	
placed_top_10_pref_count	128352.0	NaN	NaN	NaN	37.201984	
placed_pref_uni_devlet_count	128352.0	NaN	NaN	NaN	480.758913	1
placed_pref_uni_vakif_count	128352.0	NaN	NaN	NaN	75.939276	
placed_pref_uni_kktc_count	128352.0	NaN	NaN	NaN	3.700044	
placed_pref_uni_yurt_disi_count	128352.0	NaN	NaN	NaN	0.0983	

In [203…

```
missing_counts = df.isna().sum()
missing_pct = (missing_counts / len(df) * 100).sort_values(ascending=False)

missing_table = pd.DataFrame({
    "missing_count": missing_counts,
    "missing_pct": missing_pct
}).sort_values("missing_pct", ascending=False)

display(missing_table)
```

	missing_count	missing_pct
final_rank_018	79162	61.675704
final_score_018	78199	60.925424
avg_obp_018	61090	47.595674
all_tags	58082	45.252119
final_rank_012	14236	11.091374
final_score_012	12870	10.027113
avg_obp_012	4125	3.213818
placed_pref_rank_avg	3989	3.107860
faculty_name	877	0.683277
total_enrolled	0	0.000000
total_preferences	0	0.000000
placed_top_10_pref_count	0	0.000000
total_quota	0	0.000000
university_name	0	0.000000
top_9_pref_count	0	0.000000
top_3_pref_count	0	0.000000
top_1_pref_count	0	0.000000
score_type	0	0.000000
scholarship_type	0	0.000000
university_type	0	0.000000
program_code	0	0.000000
placed_top_3_pref_count	0	0.000000
placed_top_1_pref_count	0	0.000000
additional_placement	0	0.000000
placed_pref_uni_yurt_disi_count	0	0.000000
placed_pref_uni_vakif_count	0	0.000000
placed_pref_uni_kktc_count	0	0.000000
placed_pref_uni_devlet_count	0	0.000000
placed_count	0	0.000000
not_registered	0	0.000000
male	0	0.000000
is_undergraduate	0	0.000000
initial_placement_rate	0	0.000000
female	0	0.000000
department_name	0	0.000000
demand_per_quota	0	0.000000
city	0	0.000000

	missing_count	missing_pct
avg_preference_rank	0	0.000000
year	0	0.000000

Summary

The dataset contains more than 128,000 program–year observations across 39 variables, providing rich coverage of Turkish university admissions between 2019 and 2024. The structure follows a long format in which each row represents a unique degree program in a specific year. The dataset mixes numerical variables (e.g., quotas, preferences, ranks, gender counts) with categorical descriptors (e.g., university type, scholarship type, score type). This combination supports both broad descriptive analysis and fine-grained comparisons across financial categories, regions, and exam score types.

Data cleaning decisions

1. **High-missing columns**
    - Which columns are heavily missing: final_score_018 , final_rank_018 , avg_obp_018 all_tags
    - Missingness is not random: imputing would artificially distort competitiveness measures.
2. **Duplicates**
    - Each row represents a unique program–year combination; duplicates would indicate scraping or merging errors, but none were present.
    - Confirms base dataset integrity.
3. **Outliers / impossible values**
    - No removals: all values retained.
    - Zero-enrolment or zero-preference rows reflect genuinely low-demand programs.
    - Analyses use medians, IQRs, and log scales where needed to avoid distortion.
4. **Language**
    - uppercase and lowercase variations
    - turkish characters and special characters e.g. Ü, İ, Ş, Ç, Ö
    - punctuation inconsistencies e.g. hyphens

Removing Duplicates

In [204…

```
dup_rows = df.duplicated().sum()
print("Duplicate rows:", dup_rows)

# Note: If you decide duplicates are errors: df = df.drop_duplicates()
```

```
Duplicate rows: 0
```

Standardising Column Names: Justification & Impact

Ensures compatibility with Python methods. Prevents subtle bugs caused by unicode characters. Makes code reproducible and consistent across platforms.

All downstream code becomes cleaner and avoids silent failures in the future.

In [205…
```
df.columns = (
    df.columns
      .str.strip()
      .str.lower()
      .str.replace(" ", "_")
      .str.replace(r"[^a-z0-9_]", "", regex=True)
)
```

Missing Columns: Justification & Impact

Missingness is sparse for most key admissions variables, but several competitiveness metrics—especially final_score_018, final_rank_018 and related OBP measures—show substantial structural missingness. These fields appear not to have been consistently reported in all years or for all score systems. Because the missingness is systematic rather than random, imputation would distort the meaning of these metrics. It ensures ranking/score analyses remain valid while preserving full dataset for scholarship/demand analyses.

Instead, analyses involving competitiveness should either focus on the better-populated "_012" fields or filter rows accordingly. The rest of the dataset is sufficiently complete to permit reliable descriptive analysis without heavy preprocessing.

In [206…
```
df_nonmissing_rank = df[df["final_rank_012"].notna()]
```

Core Columns

In [207…
```
core_cols = [
    "program_code","year","university_name","city","university_type",
    "department_name","faculty_name","score_type","scholarship_type",
    "total_quota","total_enrolled","total_preferences","demand_per_quota",
    "initial_placement_rate","placed_count",
    "final_score_012","final_rank_012","final_score_018","final_rank_018",
    "male","female"
]

df_core = df[core_cols].copy()
df_core.head()
```

Out[207…

	program_code	year	university_name	city	university_type	department_name	faculty_n
0	106510077	2019	ABDULLAH GÜL ÜNİVERSİTESİ	KAYSERİ	devlet	Bilgisayar Mühendisliği	Mühenc Fakü
1	106510077	2020	ABDULLAH GÜL ÜNİVERSİTESİ	KAYSERİ	devlet	Bilgisayar Mühendisliği	Mühenc Fakü
2	106510077	2021	ABDULLAH GÜL ÜNİVERSİTESİ	KAYSERİ	devlet	Bilgisayar Mühendisliği	Mühenc Fakü
3	106510077	2022	ABDULLAH GÜL ÜNİVERSİTESİ	KAYSERİ	devlet	Bilgisayar Mühendisliği	Mühenc Fakü
4	106510077	2023	ABDULLAH GÜL ÜNİVERSİTESİ	KAYSERİ	devlet	Bilgisayar Mühendisliği	Mühenc Fakü

5 rows × 21 columns

Conversions for Data Types: Justification & Impacts

Ensures statistical summary functions behave properly. Prevents accidental string sorting (e.g. "100" < "20").

Critical for correlation, distribution, and regression-type analyses.

In [208…
```
numeric_cols = ["total_quota", "total_enrolled", "male", "female",
                "final_score_012","final_rank_012","final_score_018","final_rank_018"

for col in numeric_cols:
    df[col] = pd.to_numeric(df[col], errors="coerce")
```

Zero Values

Some variables contain 0 as an outcome recorded where while this is possible and important to retain in the dataset it is important to highlight and analyse, not erase.

Removing them would erase the lower tail of demand, biasing findings toward high-demand programs.

Cleaning Summary

The dataset required minimal structural cleaning but presented several non-obvious integrity issues, including structurally missing score fields, heavy-tailed distributions, and highly disjointed scholarship categories. I addressed these through a combination of filtering, categorical consolidation and ensuring consistency of data types. These decisions ensured that downstream descriptive statistics remained robust and interpretable while preserving the real competitive profile of the admissions landscape.

Descriptive statistics: numeric variables

- **demand_per_quota** (main outcome)
- **total_preferences** and **total_quota**
- **initial_placement_rate**
- **final_score_012 / final_rank_012** (as a competitiveness proxy)
- **male / female** (composition of students)

Interpretation:

- Is demand_per_quota typically low or high? Wide spread or tight?
- Are distributions skewed (long right tail)?
- What do extreme outliers represent (elite programs?)

In [209…
```
numeric_cols = ["total_quota", "total_enrolled", "male", "female",
                "final_score_012","final_rank_012","final_score_018","final_rank_018"

for col in numeric_cols:
    df[col] = pd.to_numeric(df[col], errors="coerce")
```

In [210…
```
num_df = df_core.select_dtypes(include=np.number)
```

```
desc = num_df.describe().T
desc["skew"] = num_df.skew(numeric_only=True)
desc["kurtosis"] = num_df.kurtosis(numeric_only=True)

display(desc)
```

	count	mean	std	min	25%	
program_code	128352.0	1.545565e+08	6.156625e+07	1.001100e+08	1.049502e+08	1.10
year	128352.0	2.021501e+03	1.715442e+00	2.019000e+03	2.020000e+03	2.022
total_quota	128352.0	4.748979e+01	1.619614e+02	0.000000e+00	1.300000e+01	4.000
total_enrolled	128352.0	4.455723e+01	1.547833e+02	0.000000e+00	1.000000e+01	3.400
total_preferences	128352.0	8.777020e+02	3.528987e+03	0.000000e+00	1.410000e+02	3.830
demand_per_quota	128352.0	1.874438e+01	2.171120e+01	0.000000e+00	6.600000e+00	1.400
initial_placement_rate	128352.0	8.922572e+01	2.637303e+01	0.000000e+00	1.000000e+02	1.000
placed_count	128352.0	4.438683e+01	1.548041e+02	0.000000e+00	9.000000e+00	3.300
final_score_012	115482.0	2.913568e+02	6.982880e+01	1.383854e+02	2.386268e+02	2.768
final_rank_012	114116.0	7.566999e+05	6.468492e+05	1.800000e+01	2.057695e+05	5.741
final_score_018	50153.0	2.669897e+02	4.549083e+01	1.666646e+02	2.328486e+02	2.578
final_rank_018	49190.0	1.075540e+06	5.398187e+05	2.850000e+02	6.509090e+05	1.070
male	128352.0	2.059012e+01	6.665990e+01	0.000000e+00	4.000000e+00	1.300
female	128352.0	2.379671e+01	9.788207e+01	0.000000e+00	4.000000e+00	1.300

Descriptive statistics: categorical variables

Key categories for the main question:

- scholarship_type (Ücretsiz, Burslu, Ücretli, %50 İndirimli, etc.)
- score_type (SAY, EA, SÖZ, DİL, TYT)
- university_type (devlet vs vakıf etc.)
- city

Interpretations:

- Which scholarship types dominate the dataset? Any very small categories?
- Is the distribution of score types balanced or lopsided?
- Any categories that look messy / redundant?

In [211…

```
cat_cols = df_core.select_dtypes(exclude=np.number).columns

for col in cat_cols:
    print(f"\n--- {col} ---")
    display(df_core[col].value_counts(dropna=False).head(15))
    display(df_core[col].value_counts(normalize=True, dropna=False).head(15))
```

```
--- university_name ---
```

```
university_name
İSTANBUL GELİŞİM ÜNİVERSİTESİ      2251
İSTANBUL AYDIN ÜNİVERSİTESİ        2109
İSTANBUL NİŞANTAŞI ÜNİVERSİTESİ    1960
İSTANBUL MEDİPOL ÜNİVERSİTESİ      1902
İSTANBUL BEYKENT ÜNİVERSİTESİ      1755
BAŞKENT ÜNİVERSİTESİ               1730
SELÇUK ÜNİVERSİTESİ                1534
ATATÜRK ÜNİVERSİTESİ               1469
AKDENİZ ÜNİVERSİTESİ               1382
İSTANBUL AREL ÜNİVERSİTESİ         1377
ÜSKÜDAR ÜNİVERSİTESİ               1355
DOĞU AKDENİZ ÜNİVERSİTESİ          1310
İSTANBUL OKAN ÜNİVERSİTESİ         1256
YAKIN DOĞU ÜNİVERSİTESİ            1254
BURSA ULUDAĞ ÜNİVERSİTESİ          1251
Name: count, dtype: int64
university_name
İSTANBUL GELİŞİM ÜNİVERSİTESİ      0.017538
İSTANBUL AYDIN ÜNİVERSİTESİ        0.016431
İSTANBUL NİŞANTAŞI ÜNİVERSİTESİ    0.015271
İSTANBUL MEDİPOL ÜNİVERSİTESİ      0.014819
İSTANBUL BEYKENT ÜNİVERSİTESİ      0.013673
BAŞKENT ÜNİVERSİTESİ               0.013479
SELÇUK ÜNİVERSİTESİ                0.011952
ATATÜRK ÜNİVERSİTESİ               0.011445
AKDENİZ ÜNİVERSİTESİ               0.010767
İSTANBUL AREL ÜNİVERSİTESİ         0.010728
ÜSKÜDAR ÜNİVERSİTESİ               0.010557
DOĞU AKDENİZ ÜNİVERSİTESİ          0.010206
İSTANBUL OKAN ÜNİVERSİTESİ         0.009786
YAKIN DOĞU ÜNİVERSİTESİ            0.009770
BURSA ULUDAĞ ÜNİVERSİTESİ          0.009747
Name: proportion, dtype: float64
--- city ---
city
İSTANBUL     39416
ANKARA        9889
YURTDIŞI      7623
İZMİR         4545
KONYA         3323
ANTALYA       2545
MERSİN        1698
KOCAELİ       1678
GAZİANTEP     1604
ERZURUM       1583
BURSA         1518
TRABZON       1457
ISPARTA       1420
ESKİŞEHİR     1417
KAYSERİ       1406
Name: count, dtype: int64
```

```
city
İSTANBUL     0.307093
ANKARA       0.077046
YURTDIŞI     0.059391
İZMİR        0.035410
KONYA        0.025890
ANTALYA      0.019828
MERSİN       0.013229
KOCAELİ      0.013073
GAZİANTEP    0.012497
ERZURUM      0.012333
BURSA        0.011827
TRABZON      0.011352
ISPARTA      0.011063
ESKİŞEHİR    0.011040
KAYSERİ      0.010954
Name: proportion, dtype: float64
--- university_type ---
university_type
devlet       73074
vakif        47655
kktc          6897
yurt_disi      726
Name: count, dtype: int64
university_type
devlet       0.569325
vakif        0.371284
kktc         0.053735
yurt_disi    0.005656
Name: proportion, dtype: float64
--- department_name ---
department_name
Bilgisayar Programcılığı           2388
Çocuk Gelişimi                     2296
Psikoloji                          2008
İlk ve Acil Yardım                 1982
İşletme                            1950
Bilgisayar Mühendisliği            1787
Hemşirelik                         1583
Elektrik-Elektronik Mühendisliği   1579
Muhasebe ve Vergi Uygulamaları     1549
Bankacılık ve Sigortacılık         1547
Anestezi                           1387
Tıbbi Görüntüleme Teknikleri       1383
Tıbbi Laboratuvar Teknikleri       1370
Tıp                                1359
Mimarlık                           1354
Name: count, dtype: int64
department_name
Bilgisayar Programcılığı           0.018605
Çocuk Gelişimi                     0.017888
Psikoloji                          0.015644
İlk ve Acil Yardım                 0.015442
İşletme                            0.015193
Bilgisayar Mühendisliği            0.013923
Hemşirelik                         0.012333
Elektrik-Elektronik Mühendisliği   0.012302
Muhasebe ve Vergi Uygulamaları     0.012068
Bankacılık ve Sigortacılık         0.012053
Anestezi                           0.010806
Tıbbi Görüntüleme Teknikleri       0.010775
Tıbbi Laboratuvar Teknikleri       0.010674
Tıp                                0.010588
Mimarlık                           0.010549
Name: proportion, dtype: float64
--- faculty_name ---
```

```
faculty_name
Sağlık Hizmetleri Meslek Yüksekokulu            10825
Meslek Yüksekokulu                               9502
Mühendislik Fakültesi                            6694
Sağlık Bilimleri Fakültesi                       5620
İktisadi ve İdari Bilimler Fakültesi             5436
Fen-Edebiyat Fakültesi                           4709
Eğitim Fakültesi                                 4487
Teknik Bilimler Meslek Yüksekokulu               2844
İletişim Fakültesi                               2456
Sosyal Bilimler Meslek Yüksekokulu               2401
İktisadi, İdari ve Sosyal Bilimler Fakültesi     2286
Edebiyat Fakültesi                               2213
Mühendislik ve Doğa Bilimleri Fakültesi          1998
İnsan ve Toplum Bilimleri Fakültesi              1974
Fen Fakültesi                                    1303
Name: count, dtype: int64
faculty_name
Sağlık Hizmetleri Meslek Yüksekokulu            0.084338
Meslek Yüksekokulu                              0.074031
Mühendislik Fakültesi                           0.052153
Sağlık Bilimleri Fakültesi                      0.043786
İktisadi ve İdari Bilimler Fakültesi            0.042352
Fen-Edebiyat Fakültesi                          0.036688
Eğitim Fakültesi                                0.034959
Teknik Bilimler Meslek Yüksekokulu              0.022158
İletişim Fakültesi                              0.019135
Sosyal Bilimler Meslek Yüksekokulu              0.018706
İktisadi, İdari ve Sosyal Bilimler Fakültesi    0.017810
Edebiyat Fakültesi                              0.017242
Mühendislik ve Doğa Bilimleri Fakültesi         0.015567
İnsan ve Toplum Bilimleri Fakültesi             0.015380
Fen Fakültesi                                   0.010152
Name: proportion, dtype: float64
--- score_type ---
score_type
TYT    60030
SAY    29916
EA     22245
SÖZ    12288
DİL     3873
Name: count, dtype: int64
score_type
TYT    0.467698
SAY    0.233078
EA     0.173312
SÖZ    0.095737
DİL    0.030175
Name: proportion, dtype: float64
--- scholarship_type ---
scholarship_type
Ücretsiz         63799
Burslu           21840
%50 İndirimli    18256
İÖ-Ücretli       11455
Ücretli           6930
%25 İndirimli     2385
%75 İndirimli     2006
UÖ-Ücretli         877
AÖ-Ücretli         791
UE-Ücretli          13
Name: count, dtype: int64
```

```
scholarship_type
Ücretsiz          0.497063
Burslu            0.170157
%50 İndirimli     0.142234
İÖ-Ücretli        0.089247
Ücretli           0.053992
%25 İndirimli     0.018582
%75 İndirimli     0.015629
UÖ-Ücretli        0.006833
AÖ-Ücretli        0.006163
UE-Ücretli        0.000101
Name: proportion, dtype: float64
```

Scholarship type as the biggest structural driver of demand

Main scholarship groups (counts + demand)

Ücretsiz/Non-paid (n≈63,799): mean 19.43, median 15.7

Burslu/Scholarship (n≈21,840): mean 29.66, median 24.5

%50 İndirimli/Partial Discount (n≈18,256): mean 7.94, median 5.9

Ücretli/Paid (n≈6,930): mean 12.89, median 5.6

%25, %75 İndirimli/Other amount of scholarships: smaller therefore noisier

There's a very clear price-sensitivity gradient: full scholarship > free public > paid private > partial discount on revealed preference (demand per quota).

The fact that %50 İndirimli is lower than Ücretli in median terms is telling: partial scholarships appear insufficient to shift student choice as strongly as either full support or free public places.

%75 İndirimli has a high mean but low median + huge stdso likely a mixture of a few hyper-competitive discounted programs and many low-demand ones (classic case of outlier-driven mean).

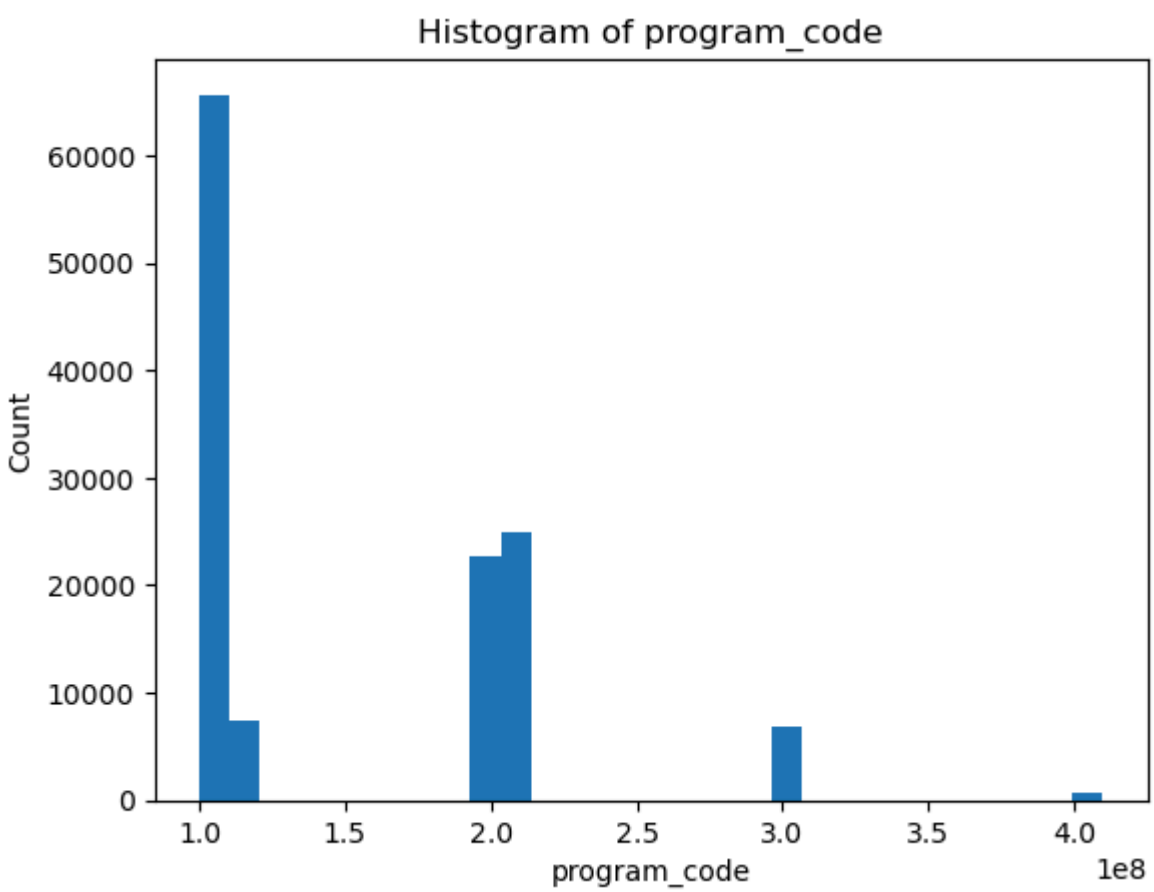
Interpretation

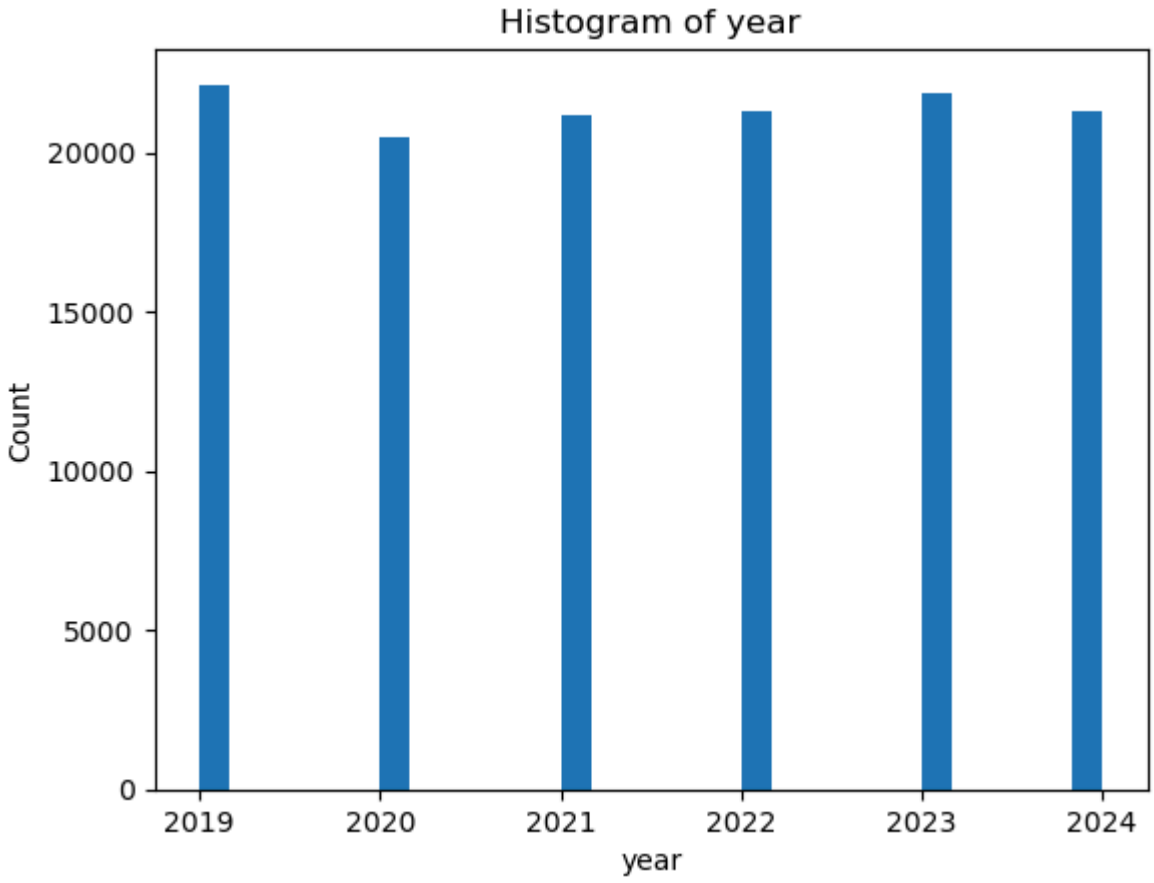
Scholarship differences aren't just "statistically significant"; they're huge.

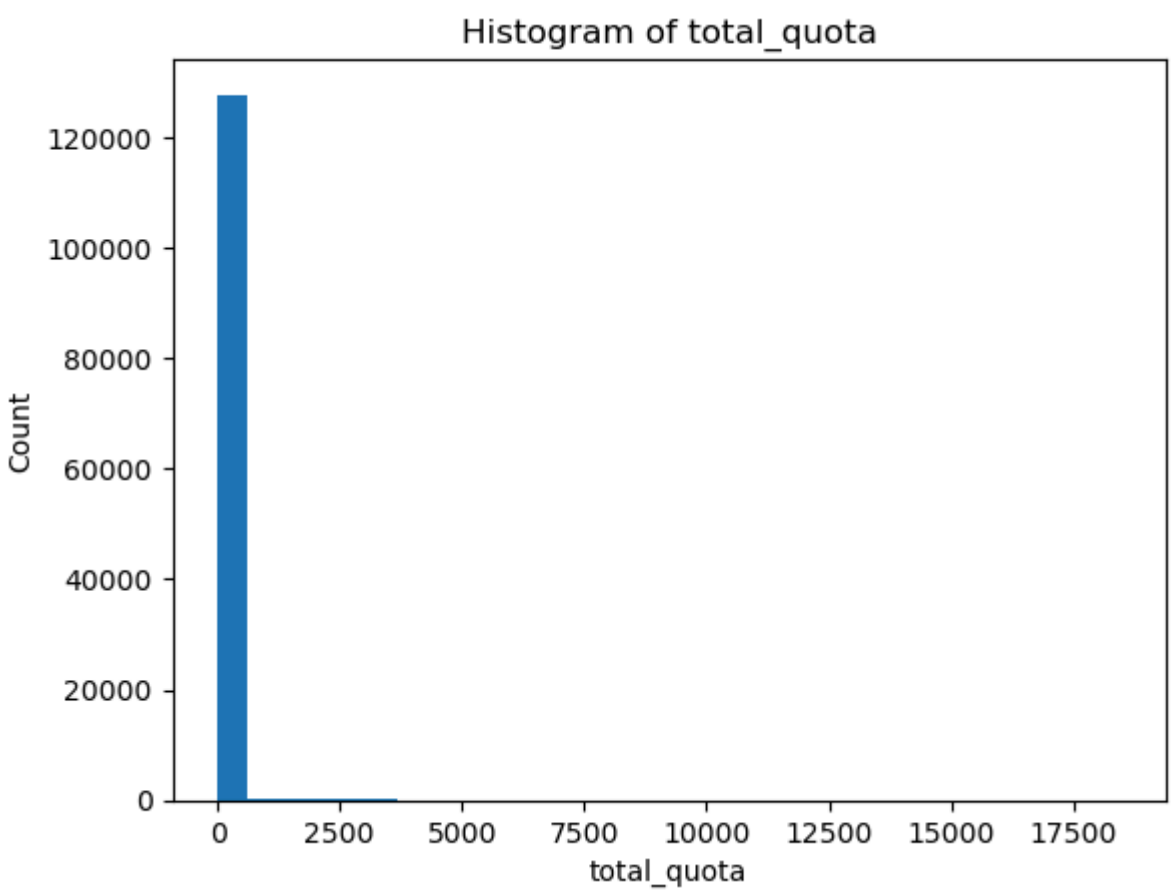
The near-zero effect between %50 İndirimli and Ücretli suggests those two segments attract similar demand intensity once we account for skew.

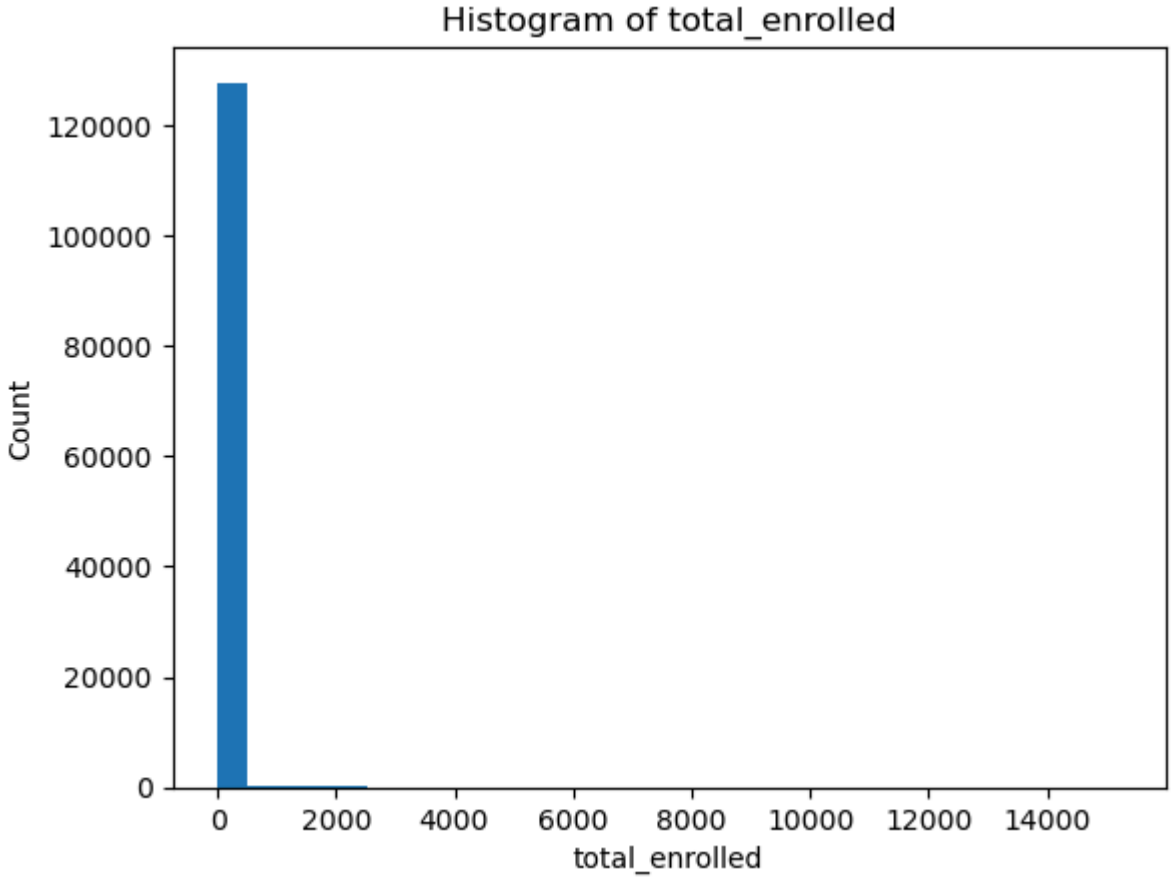
In [212…

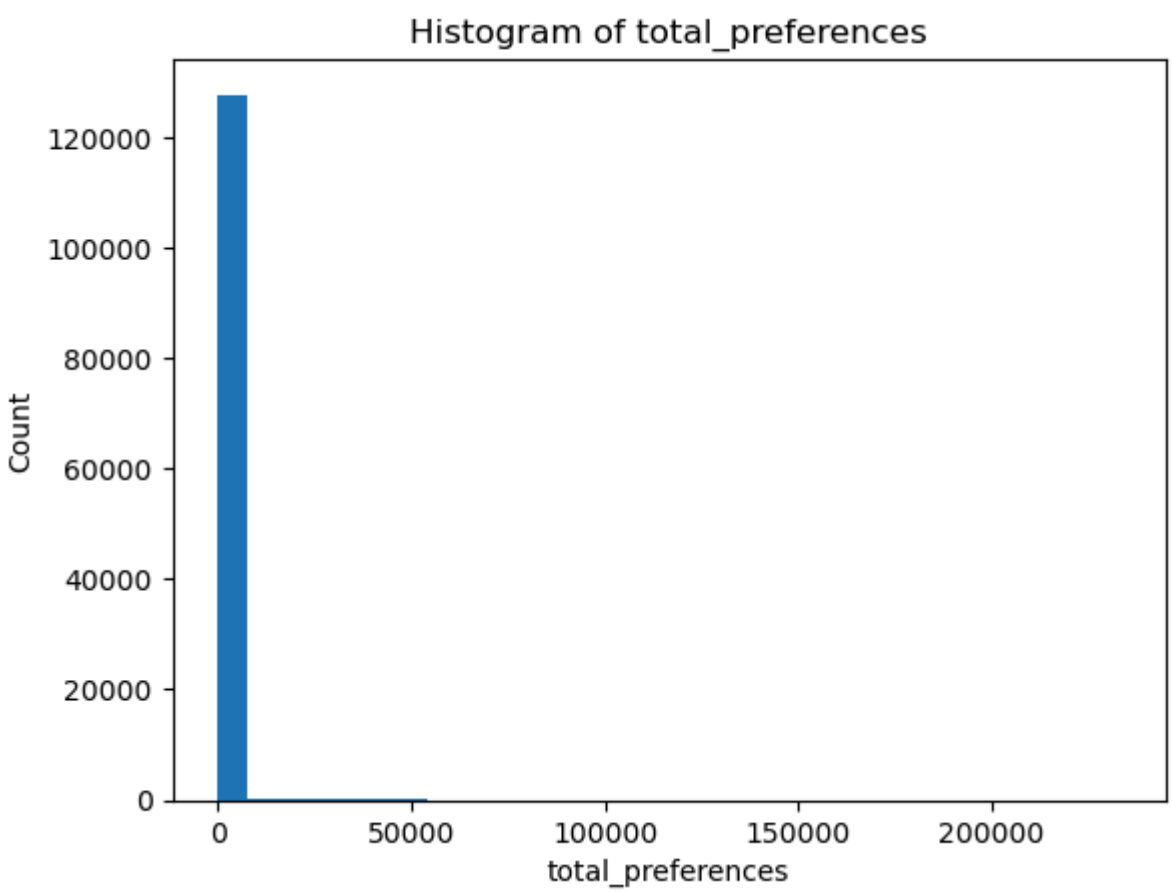
```python
for col in num_df.columns:
    plt.figure()
    plt.hist(num_df[col].dropna(), bins=30)
    plt.title(f"Histogram of {col}")
    plt.xlabel(col)
    plt.ylabel("Count")
    plt.show()
```

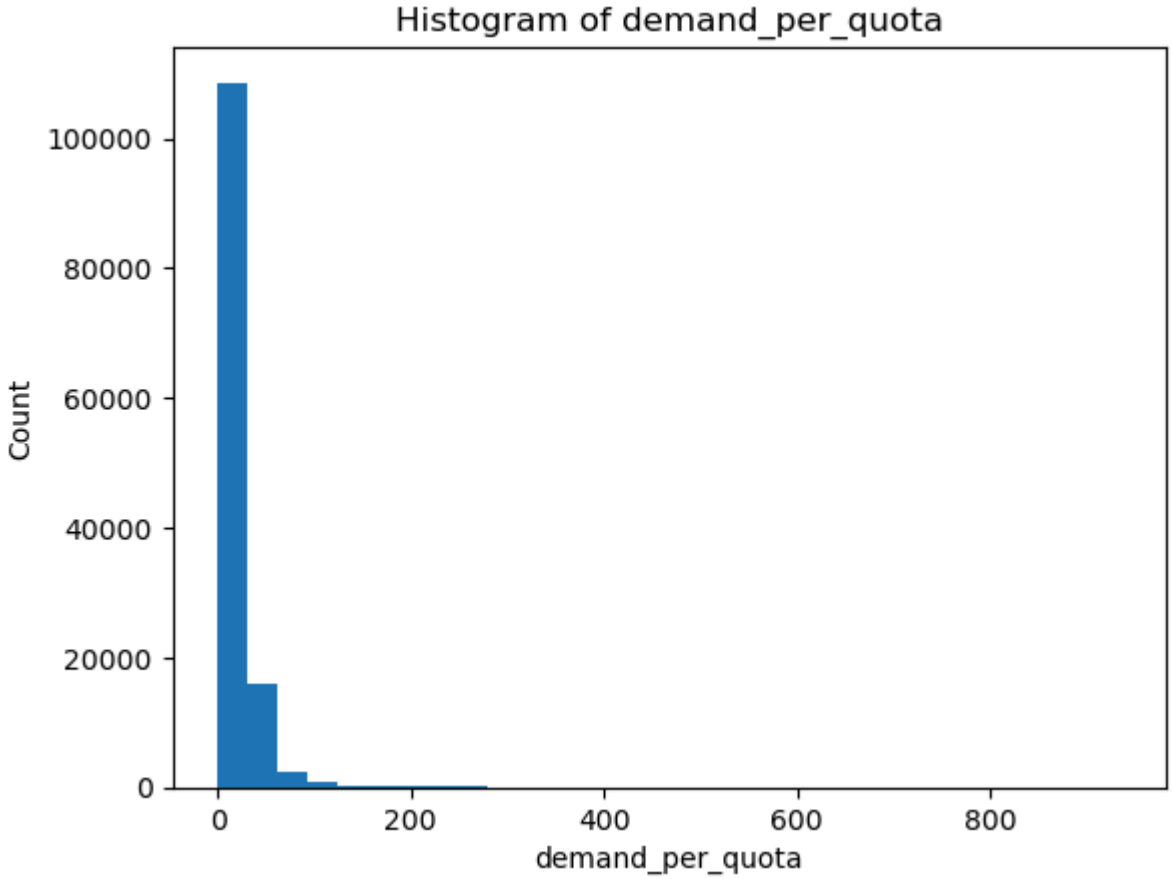

Histogram of program_code
Count
60000
50000
40000
30000
20000
10000
0
1.0
1.5
2.0
2.5
3.0
3.5
4.0
program_code
1e8

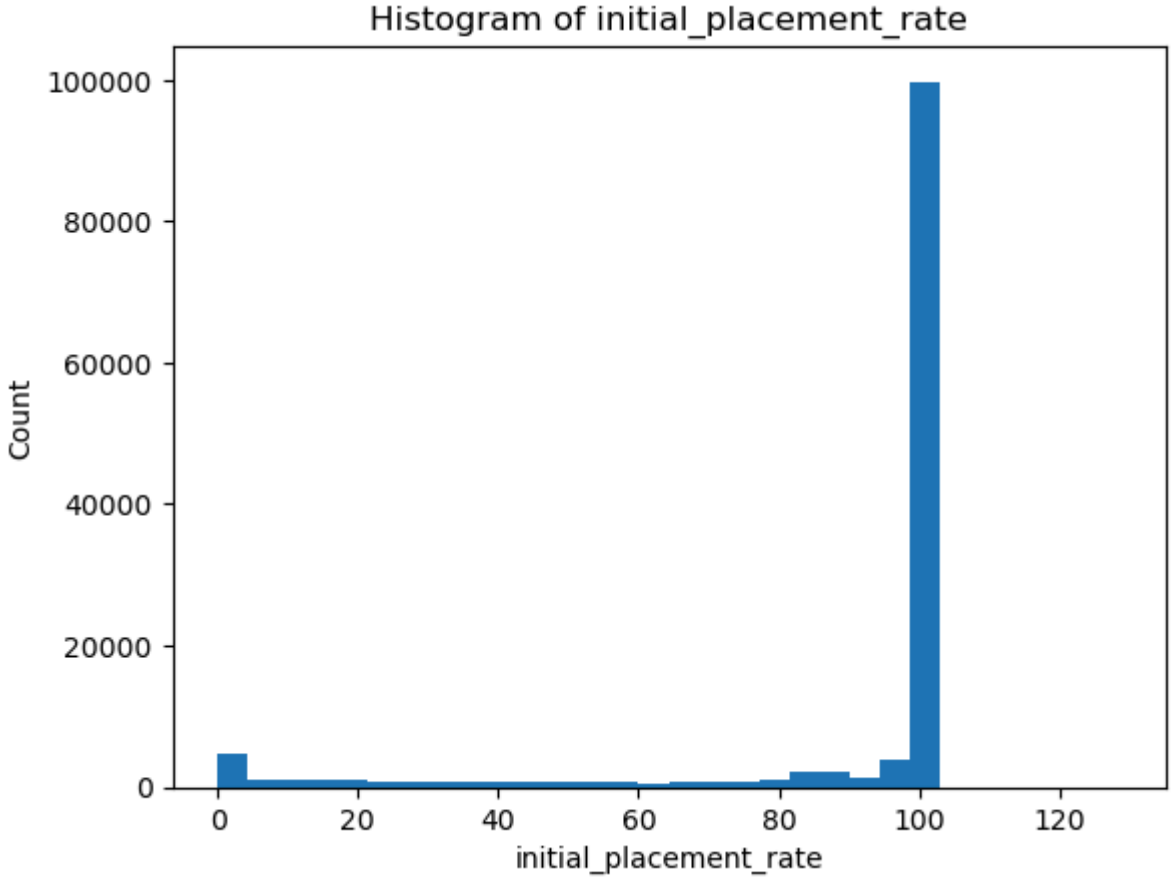

Histogram of year
Count
20000
15000
10000
5000
0
2019
2020
2021
2022
2023
2024
year

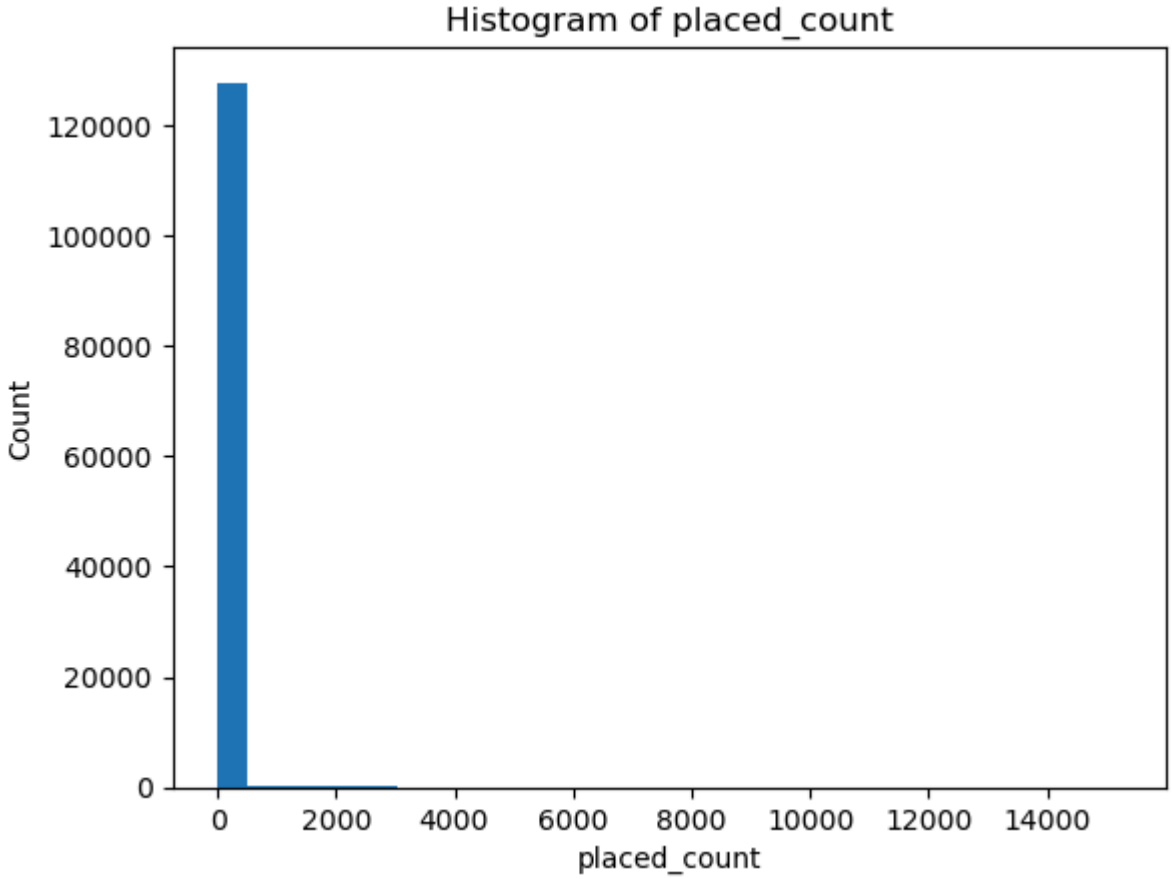

Histogram of total_quota
Count
120000
100000
80000
60000
40000
20000
0
0
2500
5000
7500
10000
12500
15000
17500
total_quota

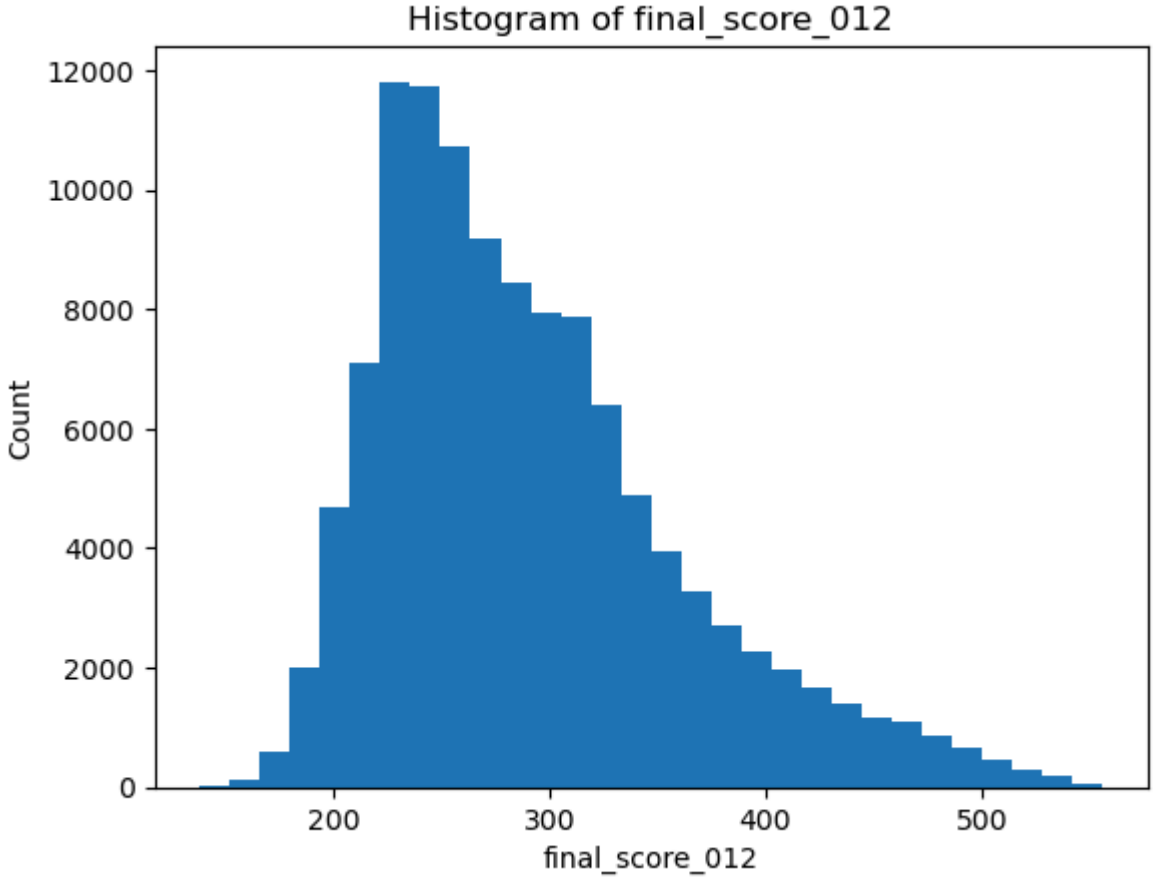

Histogram of total_enrolled
Count
120000
100000
80000
60000
40000
20000
0
0
2000
4000
6000
8000
10000
12000
14000
total_enrolled

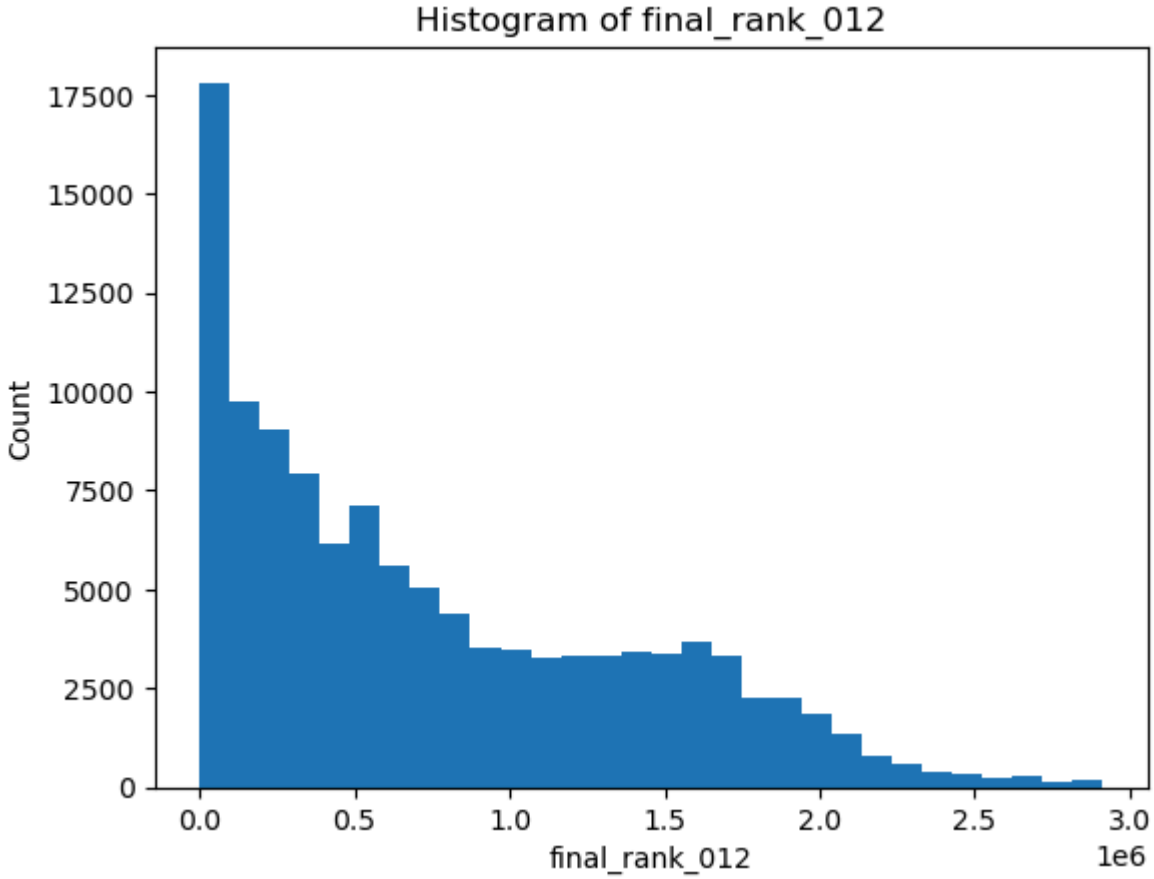

Histogram of total_preferences
Count
120000
100000
80000
60000
40000
20000
0
0
50000
100000
150000
200000
total_preferences

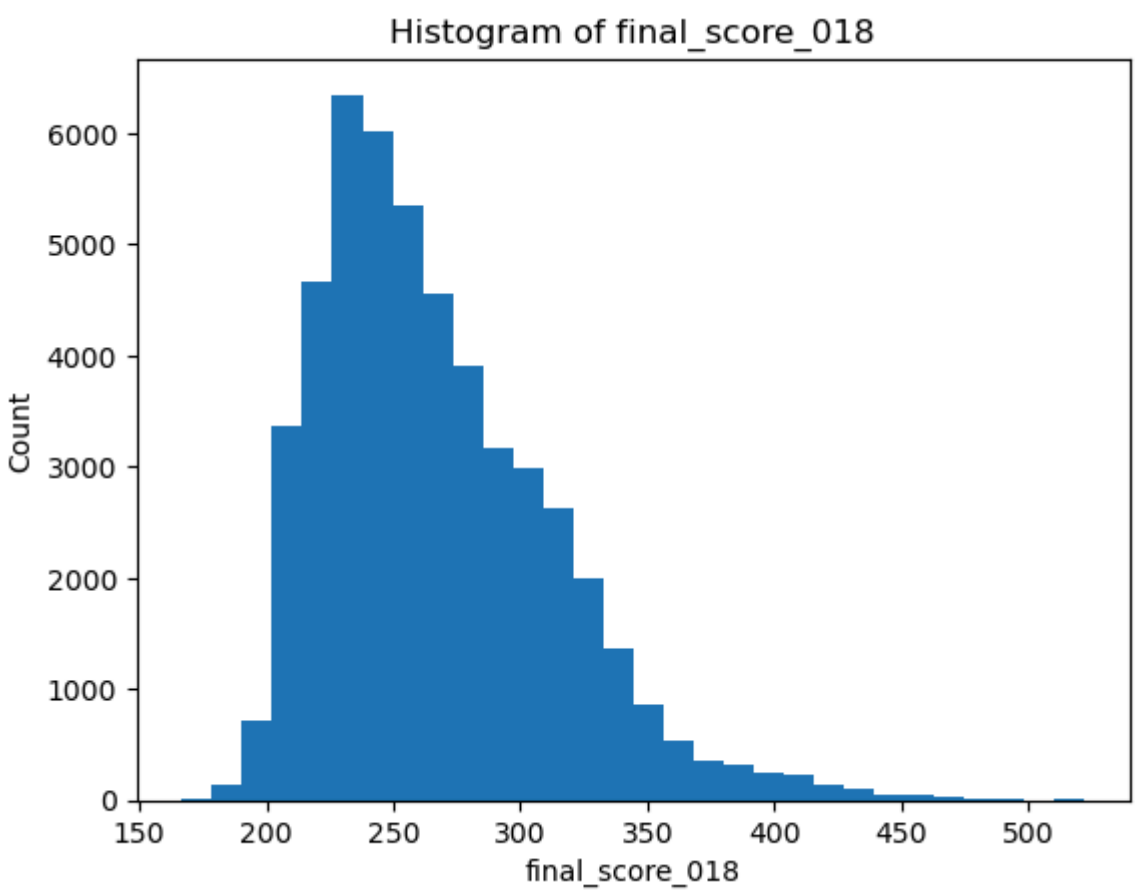

Histogram of demand_per_quota
Count
100000
80000
60000
40000
20000
0
0
200
400
600
800
demand_per_quota

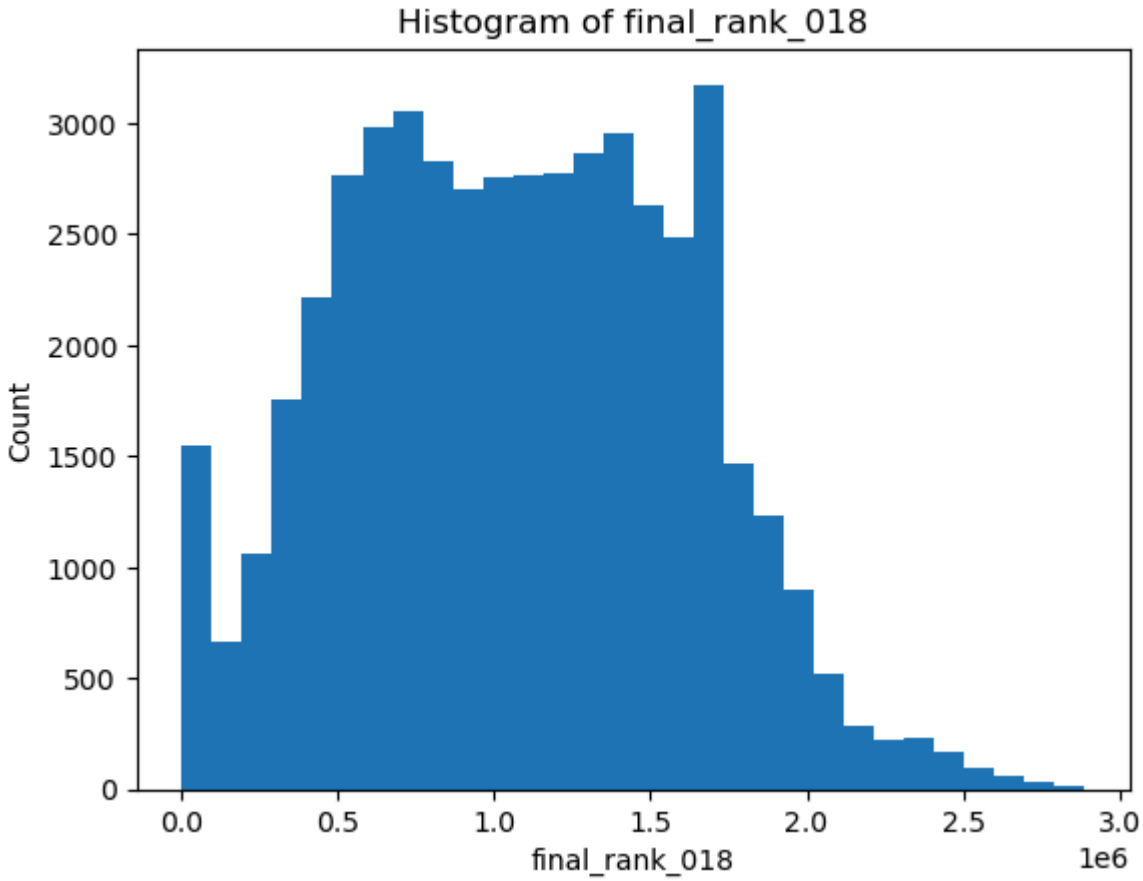

Histogram of initial_placement_rate
Count
initial_placement_rate
0
20000
40000
60000
80000
100000
0
20
40
60
80
100
120

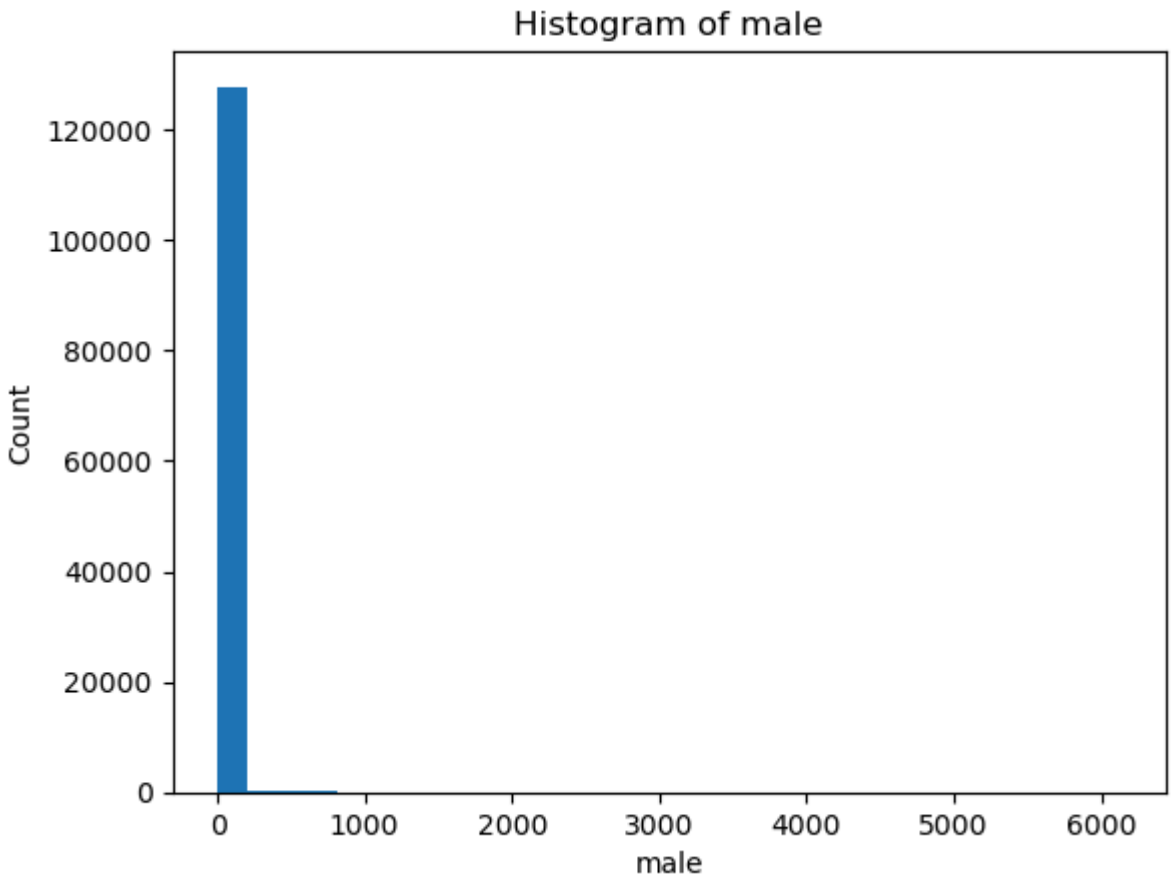

Histogram of placed_count
Count
placed_count
0
20000
40000
60000
80000
100000
120000
0
2000
4000
6000
8000
10000
12000
14000

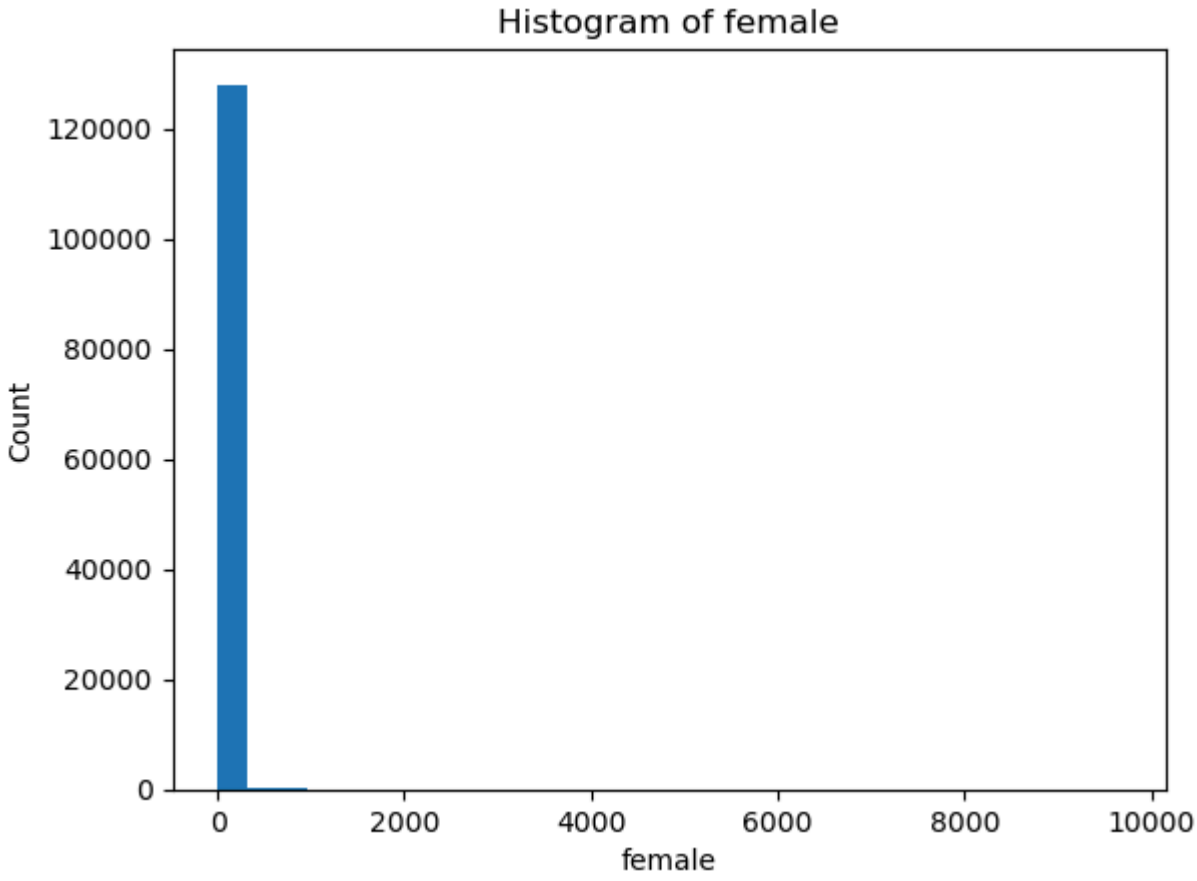

Histogram of final_score_012
Count
final_score_012
0
2000
4000
6000
8000
10000
12000
200
300
400
500


Histogram of final_rank_012
Count
final_rank_012
1e6
0
2500
5000
7500
10000
12500
15000
17500
0.0
0.5
1.0
1.5
2.0
2.5
3.0


Histogram of final_score_018
Count
6000
5000
4000
3000
2000
1000
0
150
200
250
300
350
400
450
500
final_score_018

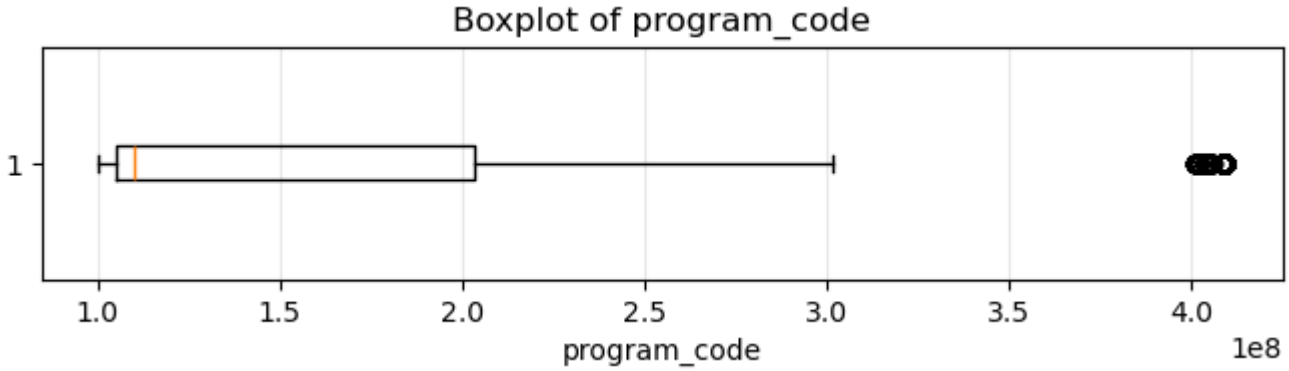

Histogram of final_rank_018
Count
3000
2500
2000
1500
1000
500
0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
final_rank_018
1e6

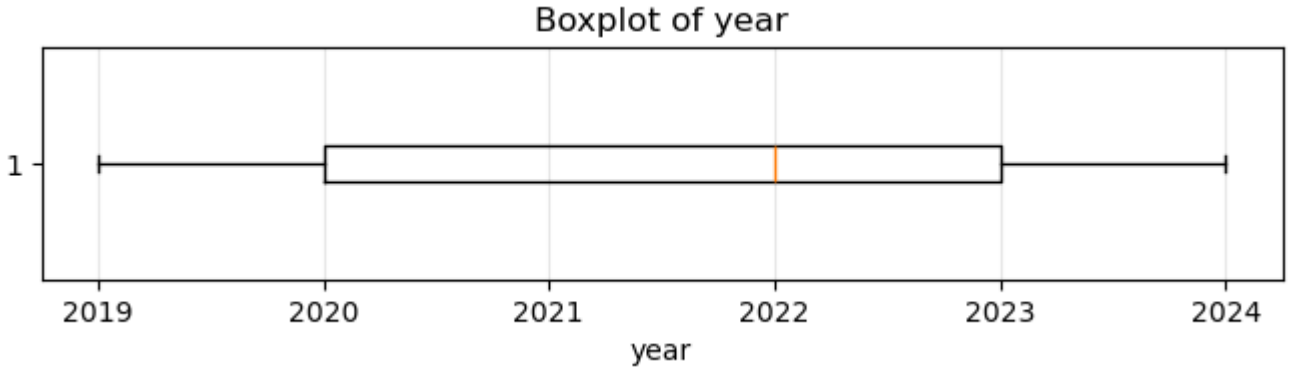


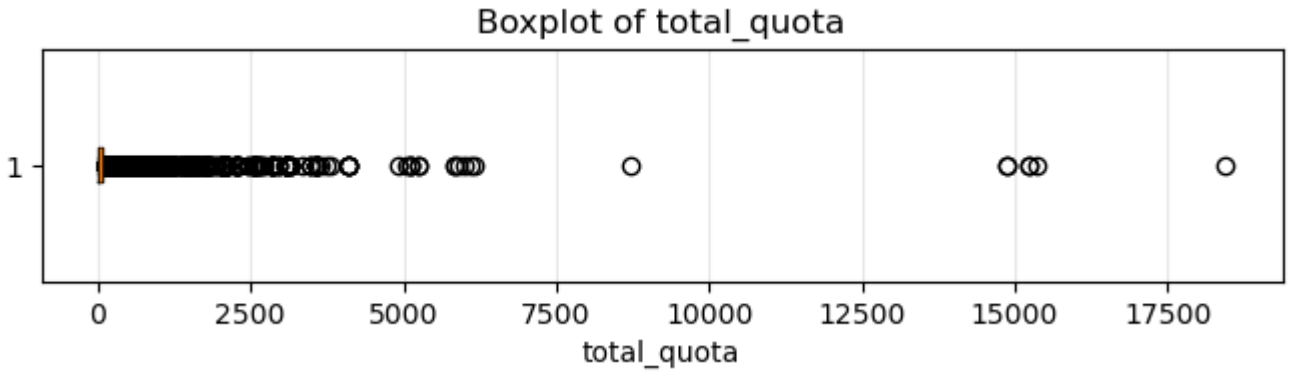


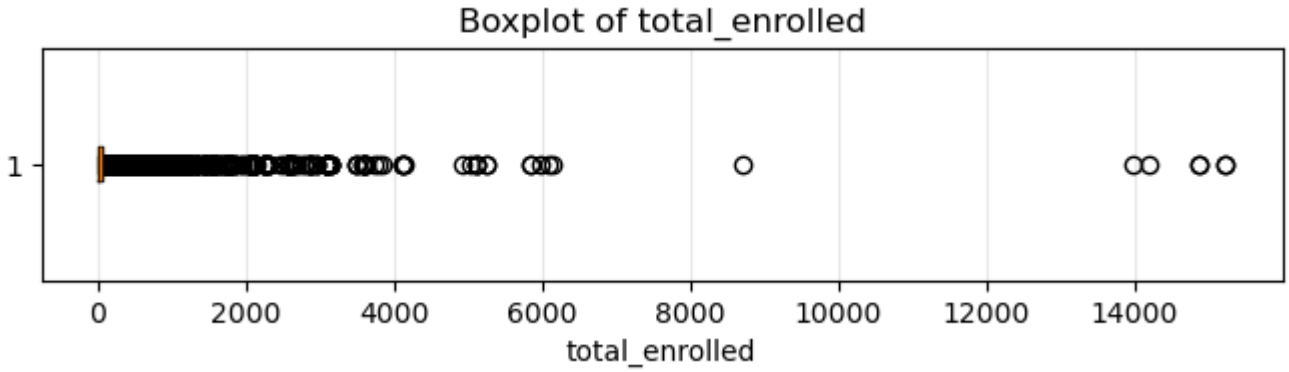
In [213…

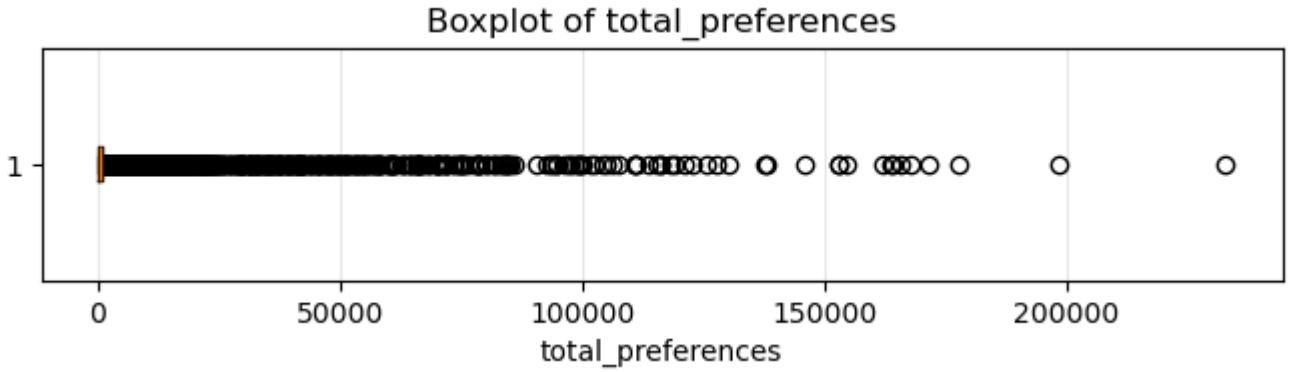
```python
for col in num_df.columns:
    plt.figure(figsize=(8, 1.5))
    plt.boxplot(num_df[col].dropna(), vert=False, showfliers=True)
    plt.title(f"Boxplot of {col}")
    plt.xlabel(col)
    plt.grid(axis="x", alpha=0.3)
    plt.show()
```

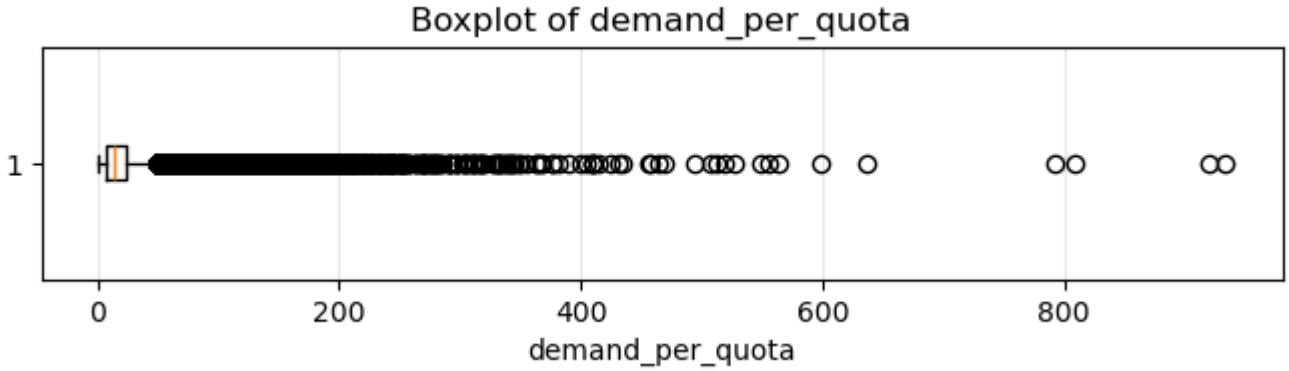

Boxplot of program_code
1
1.0
1.5
2.0
2.5
3.0
3.5
4.0
program_code
1e8

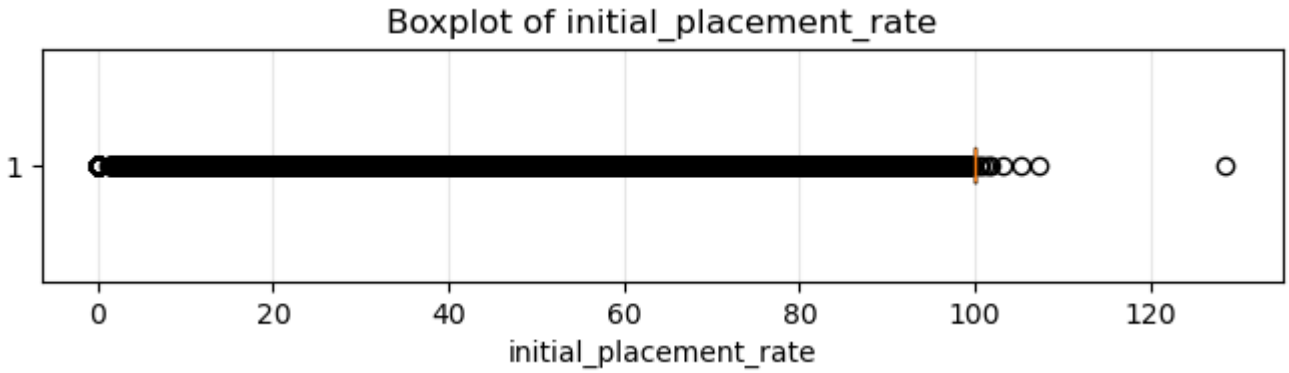

Boxplot of year
1
2019
2020
2021
2022
2023
2024
year

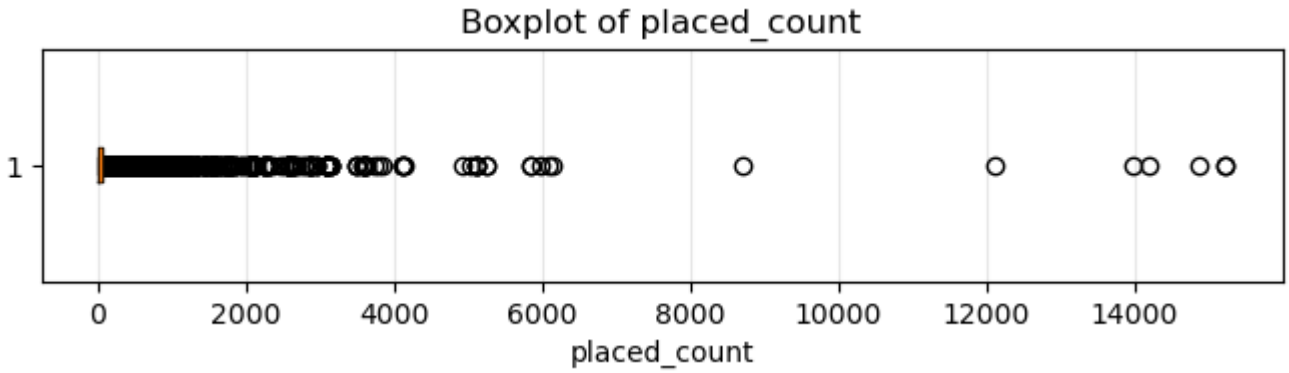

Boxplot of total_quota
1
0
2500
5000
7500
10000
12500
15000
17500
total_quota

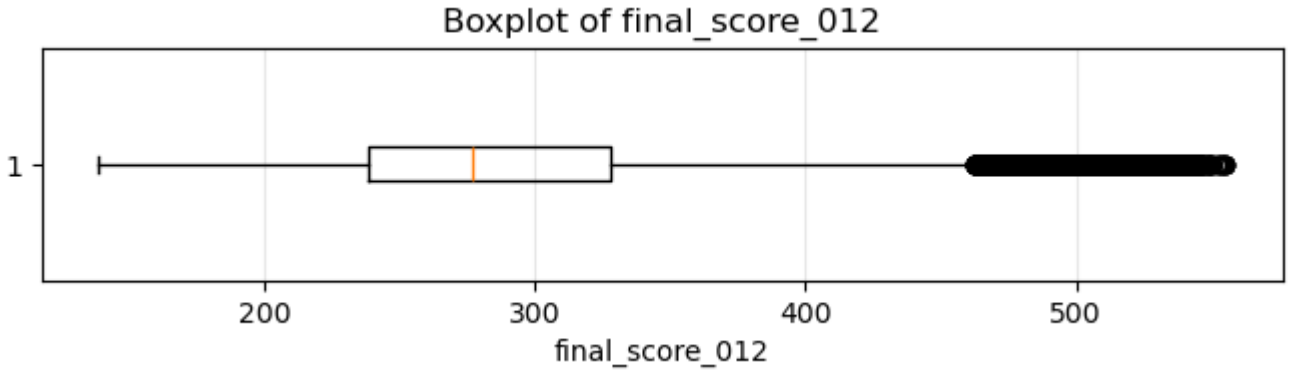

Boxplot of total_enrolled
1
0
2000
4000
6000
8000
10000
12000
14000
total_enrolled

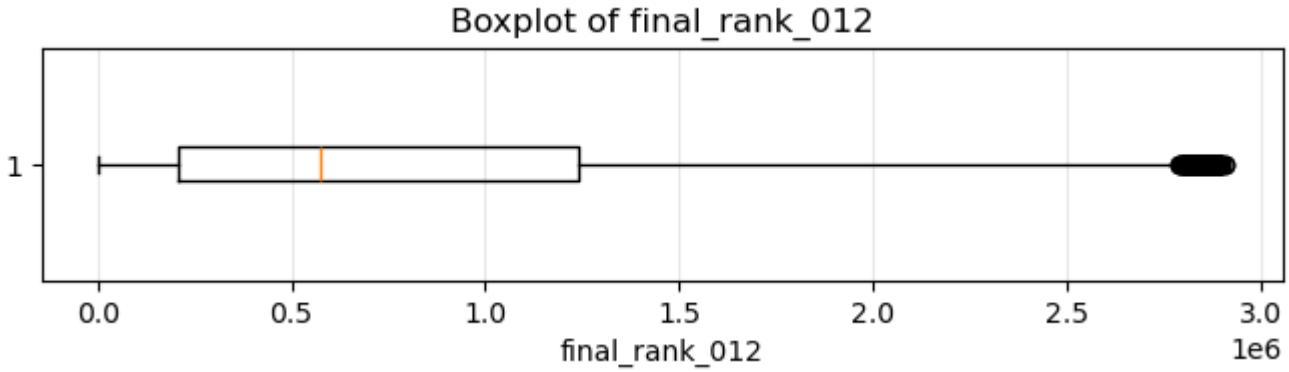

Boxplot of total_preferences
1
0
50000
100000
150000
200000
total_preferences

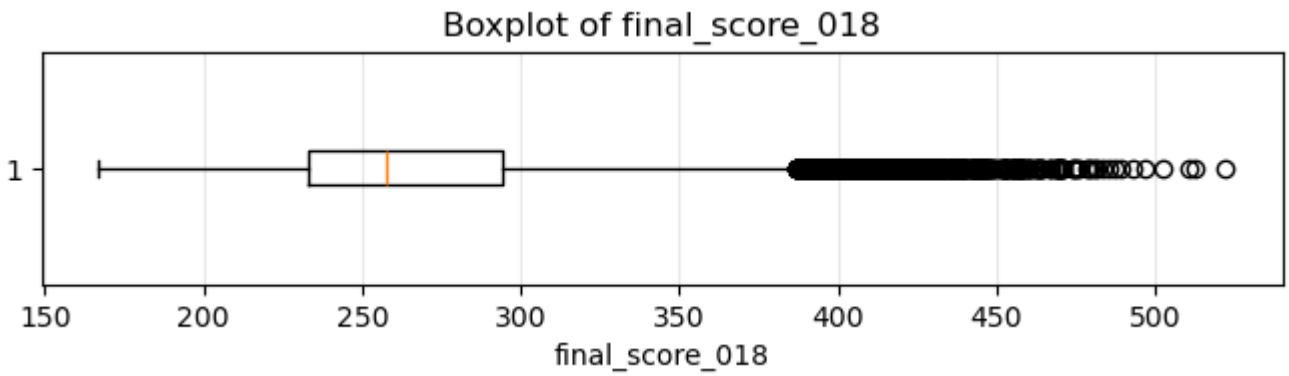

Boxplot of demand_per_quota
1
0
200
400
600
800
demand_per_quota

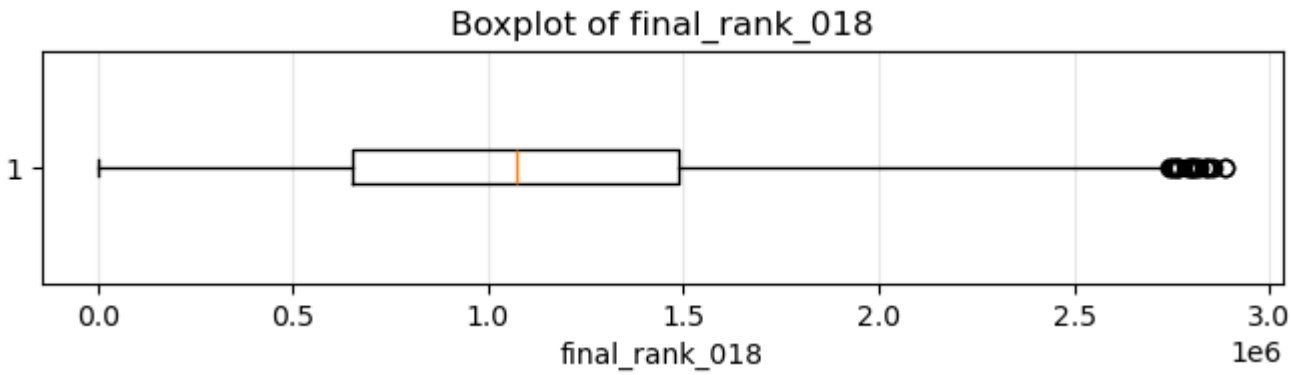

Boxplot of initial_placement_rate
1
0
20
40
60
80
100
120
initial_placement_rate

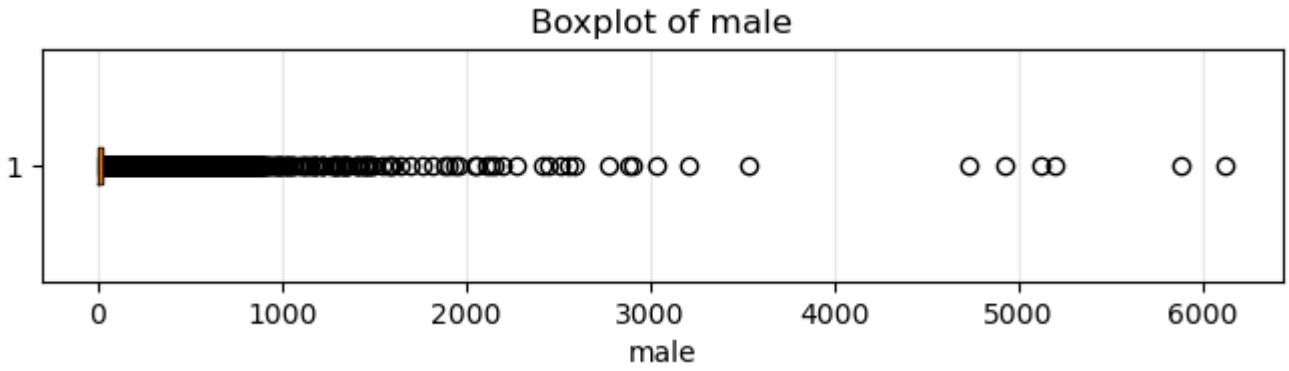

Boxplot of placed_count
1
0
2000
4000
6000
8000
10000
12000
14000
placed_count

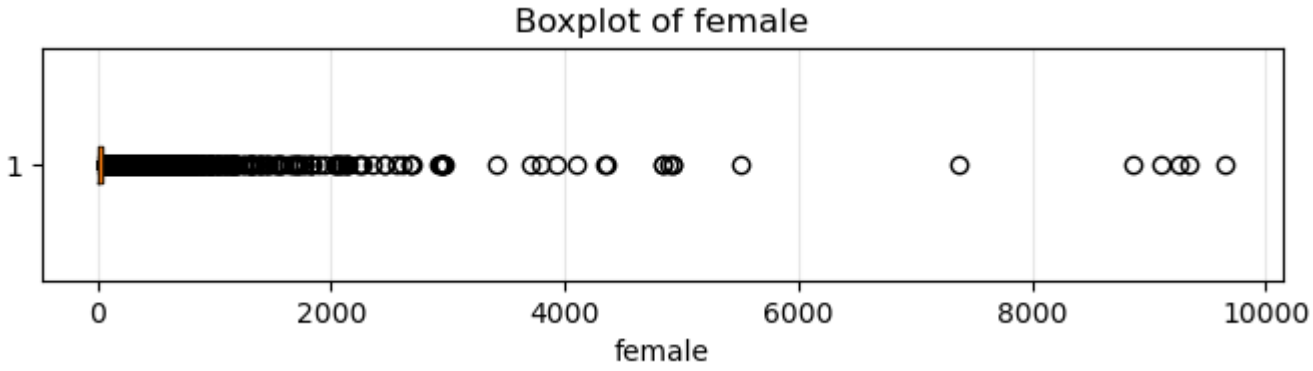

Boxplot of final_score_012
1
200
300
400
500
final_score_012


Boxplot of final_rank_012
1
0.0
0.5
1.0
1.5
2.0
2.5
3.0
final_rank_012
1e6



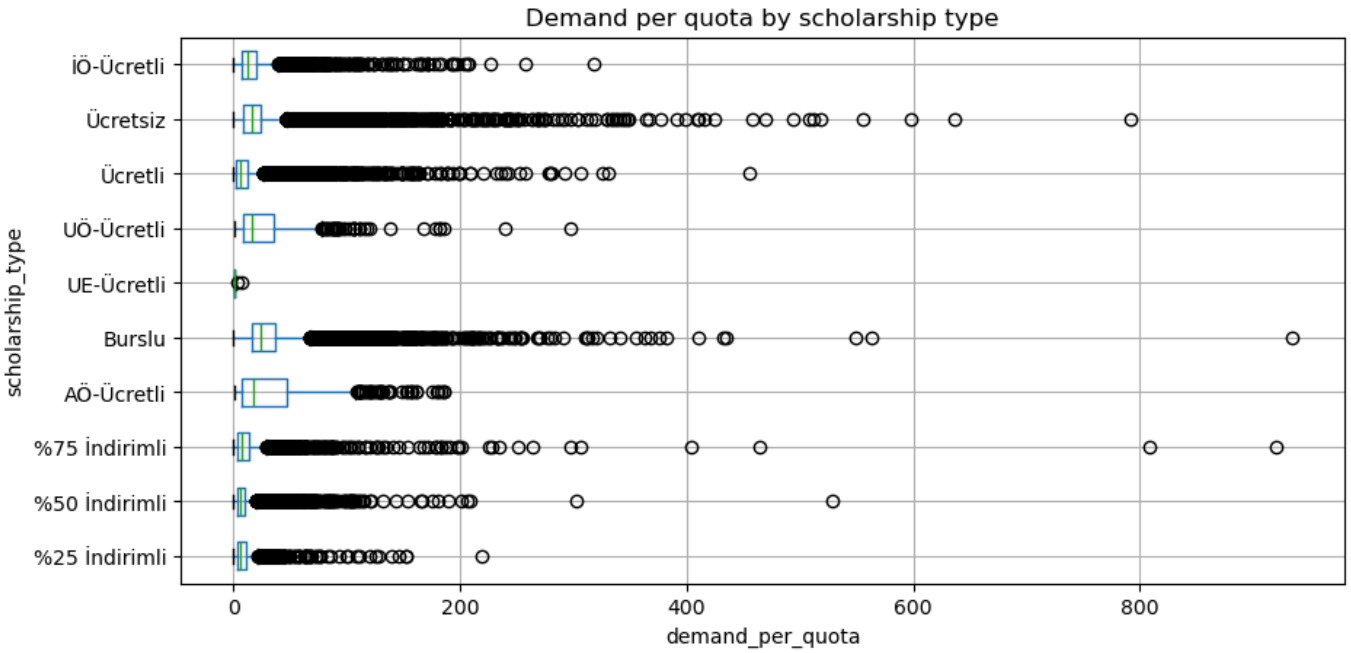






In [214…

```
## combined
df.boxplot(column="demand_per_quota", by="scholarship_type", figsize=(10,5), vert=Fal
plt.title("Demand per quota by scholarship type")
plt.suptitle("")
plt.xlabel("demand_per_quota")
plt.show()
```



Boxplot Summary

Heavy skews Most numeric variables in this dataset exhibit substantial right-skew, with long upper tails. This indicates that a small subset of programs attract disproportionately high demand or extremely favourable ranks. The presence of extreme outliers is a feature of the admissions landscape rather than data errors.

Median as < Mean For demand_per_quota, the central box is compressed near low values. Outliers extend far to the right. The median demand per quota is far lower than the upper range, reflecting that intense competition is limited to a small number of high-status programs. Most programs operate in a moderate-demand range, while a few elite departments drive the upper extremes.

Gender Boxplots for male and female counts have lower variance, tighter IQR, fewer extreme outliers. Gender distributions are more stable than demand or preferences, suggesting that although program choices differ by field, the magnitude of variation is less extreme compared with competitiveness metrics.

Quotas Some programs have very large quotas (major faculties). Others have extremely small quotas (specialist departments). Preferences vary massively.

Extreme Outliers Many variables produce single points far beyond the whiskers. These represent Turkey's most competitive programs, flagship universities, departments with small quotas and high prestige. Outliers likely correspond to elite programs rather than noise. Their presence supports the use of robust statistics (median, IQR) to avoid misleading conclusions driven by a few extreme programs.

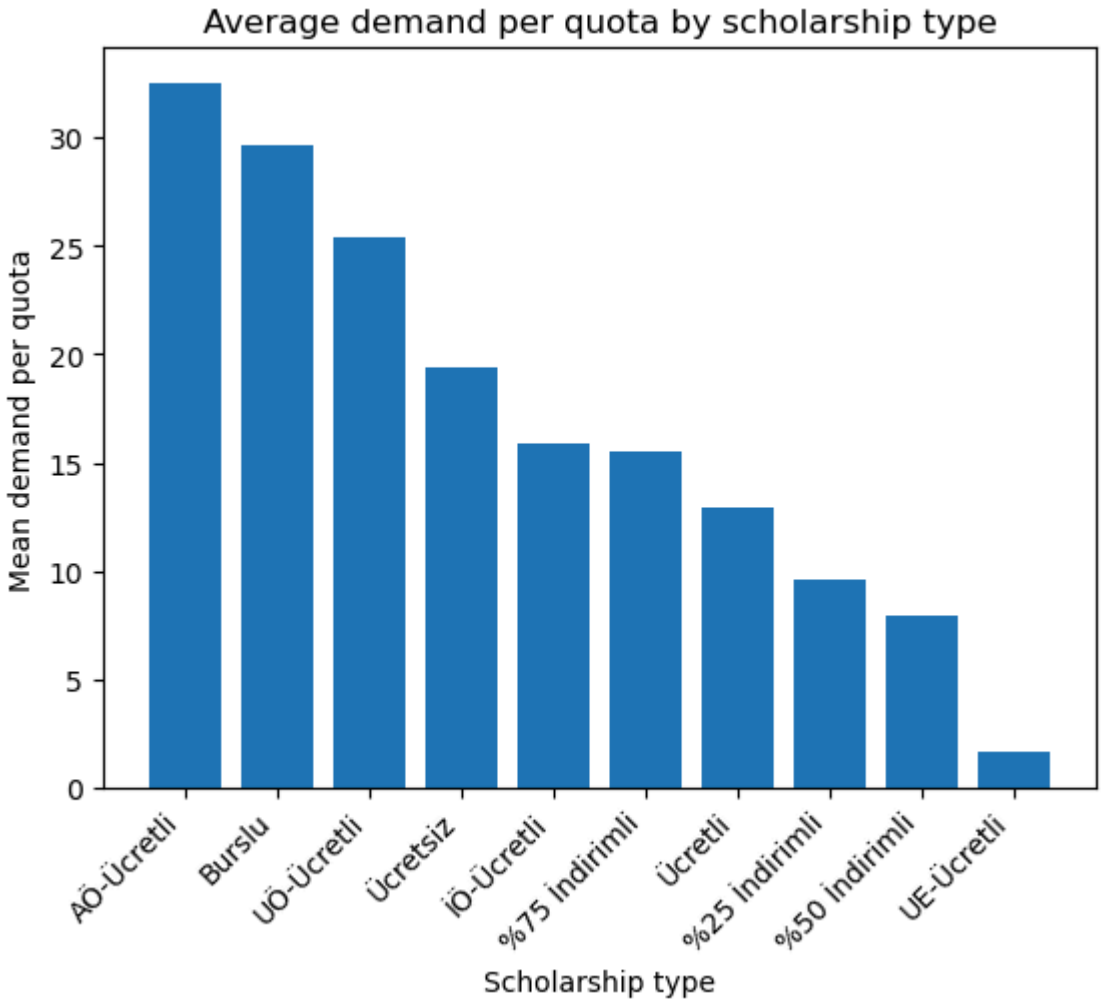
Distribution insights

Visual inspections of distributions confirm strong right-skew across key metrics. For example, demand_per_quota is concentrated near lower values but includes a long tail of extremely high-demand programs, likely representing prestigious universities or fields with national shortages (such as elite engineering or health sciences). Similar patterns in total_preferences support this view: a small number of programs attract disproportionate attention. These shapes suggest that median and IQR provide more representative summaries than means, and that log-transformations may uncover clearer structure in later analytical steps.

In [215…

```
demand_by_sch = (
    df_core.groupby("scholarship_type")["demand_per_quota"]
    .agg(["mean","median","std","count"])
    .sort_values("mean", ascending=False)
)

display(demand_by_sch)

plt.figure()
plt.bar(demand_by_sch.index, demand_by_sch["mean"])
plt.title("Average demand per quota by scholarship type")
plt.xlabel("Scholarship type")
plt.ylabel("Mean demand per quota")
plt.xticks(rotation=45, ha="right")
plt.show()
```

scholarship_type	mean	median	std	count
AÖ-Ücretli	32.495828	17.2	36.191625	791
Burslu	29.659693	24.5	25.972252	21840
UÖ-Ücretli	25.396237	16.0	27.574091	877
Ücretsiz	19.430701	15.7	20.112111	63799
İÖ-Ücretli	15.906478	11.9	15.979910	11455
%75 İndirimli	15.486391	6.7	40.041231	2006
Ücretli	12.888066	5.6	24.844928	6930
%25 İndirimli	9.626122	6.4	13.278592	2385
%50 İndirimli	7.937489	5.9	10.256016	18256
UE-Ücretli	1.676923	1.0	1.744589	13



Burslu (full scholarship) programs show the highest average demand, indicating that students strongly prioritise financial support when ranking their university choices. These programs attract far more interest per available seat than any other category.

Ücretsiz (free public university) programs also display high demand, reinforcing the importance of affordability. Even without additional scholarship incentives, free programs remain highly competitive.

Ücretli (paid) and partial discount categories (%50, %25, %75) show substantially lower demand, suggesting that students are highly price-sensitive. Financial barriers reduce the intensity of competition for these programs.

Smaller scholarship categories show more variability, but the overall trend is clear: the more financially accessible the program, the higher the demand per quota.

The steep gradient between Burslu/Ücretsiz and discount/paid categories highlights that financial support is a major structural driver of student preference behaviour, more influential than exam score type or program characteristics in explaining variation in competitiveness.

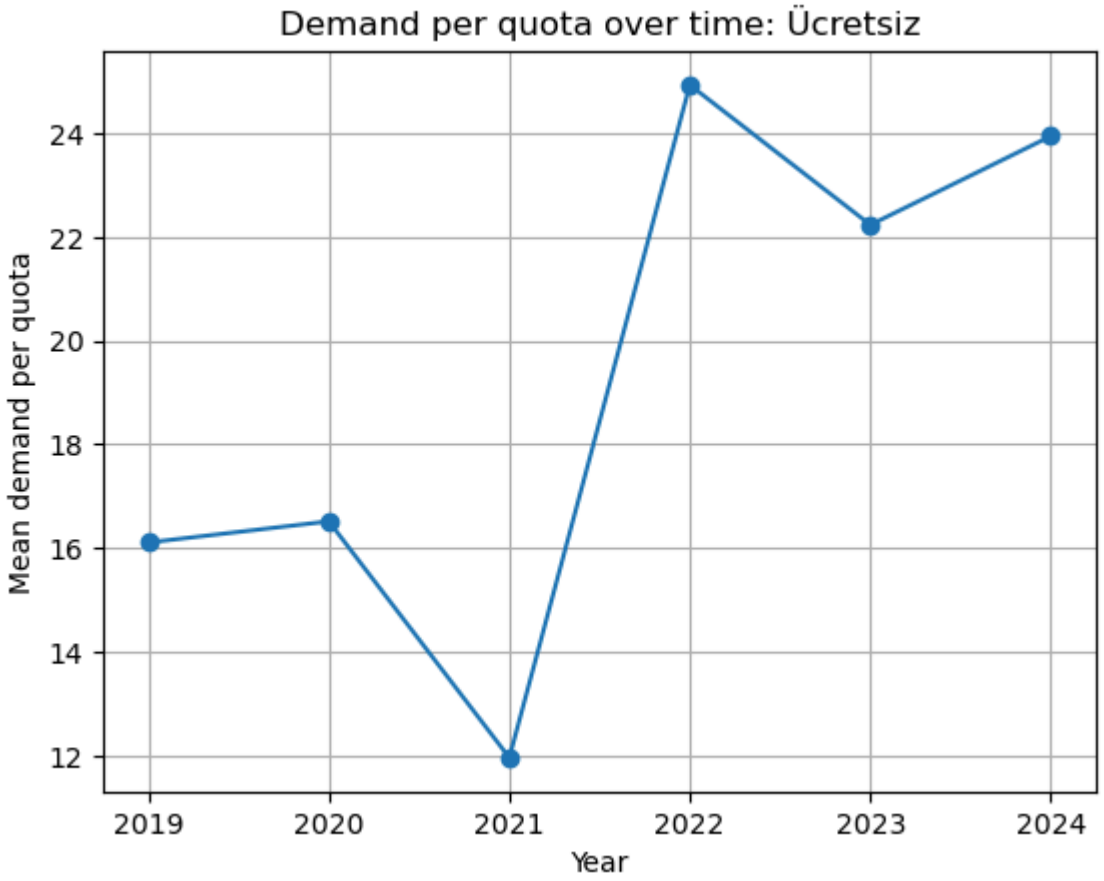
Scholarship type and demand

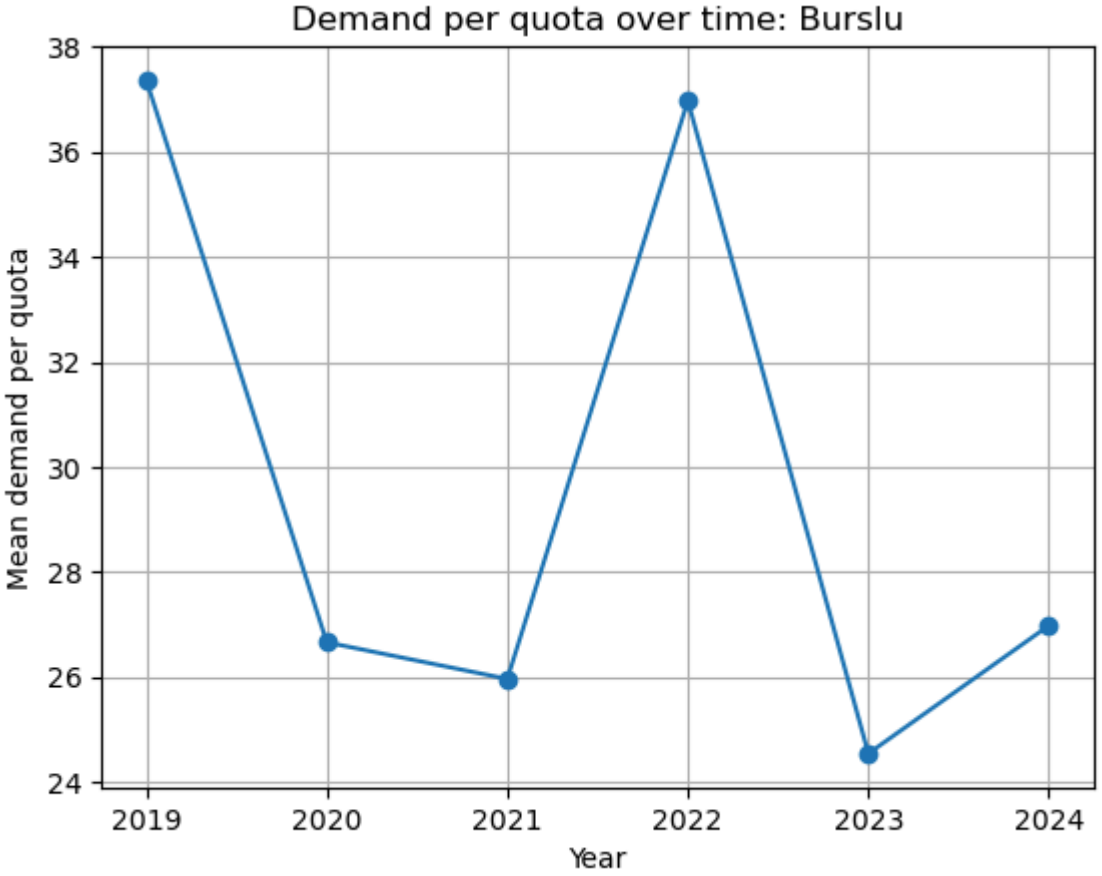
Interpretation prompts:

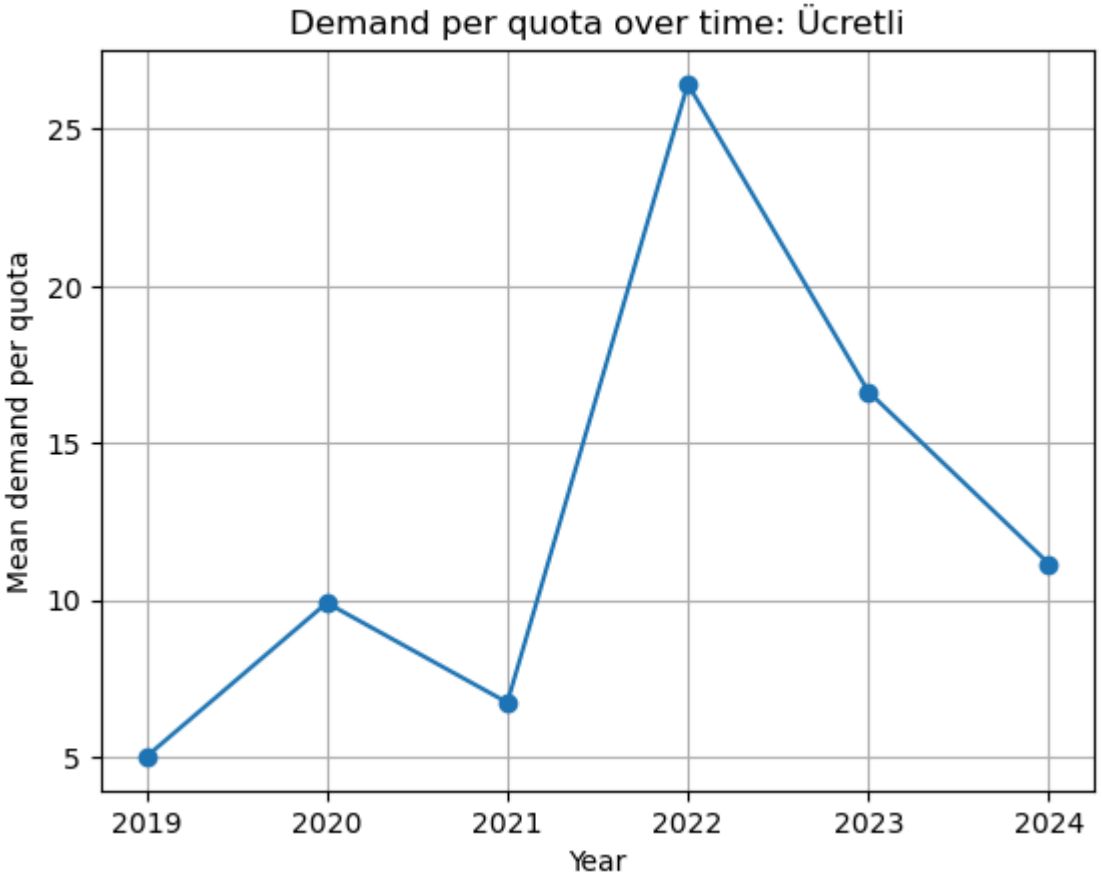
- Which scholarship categories have systematically higher demand per quota?
- Are some categories based on very small sample sizes?
- What might explain higher demand for “Burslu” vs “Ücretli” vs “Ücretsiz” programs?
- Do any results look counterintuitive or potentially data-driven (e.g., quota denominators)?

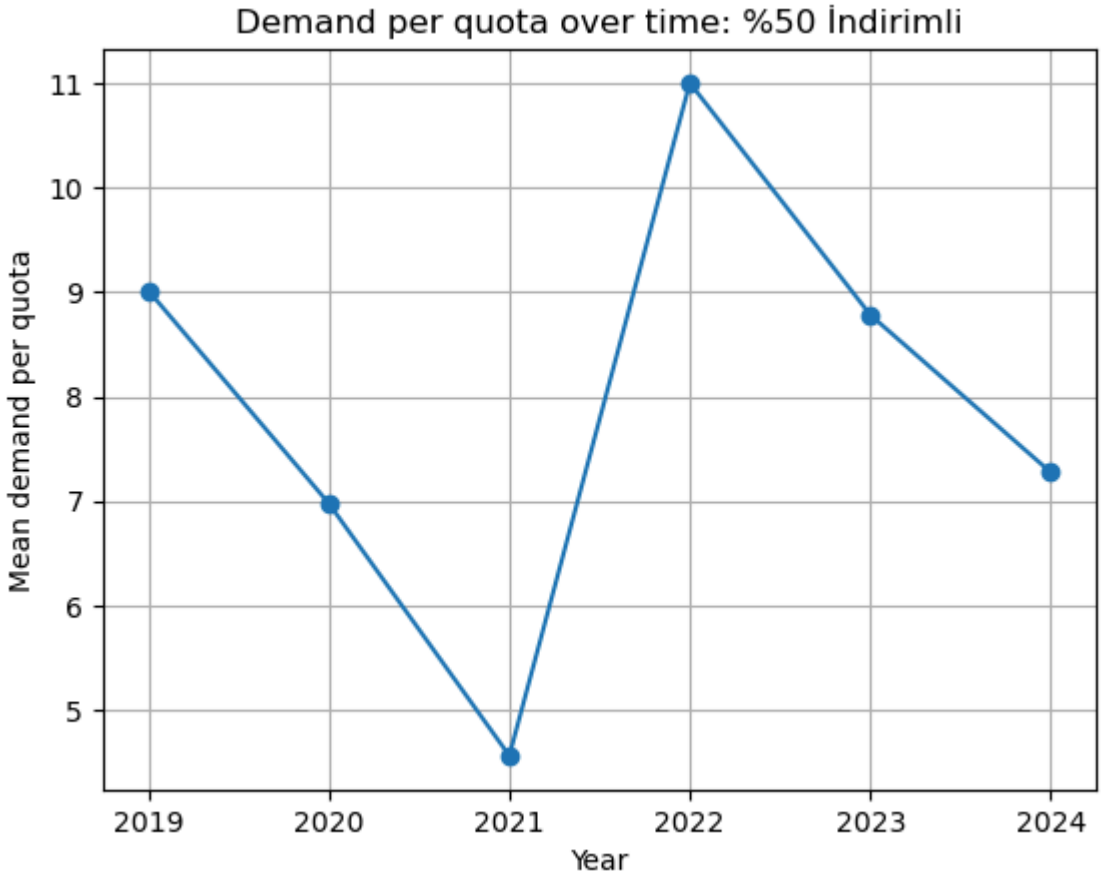
In [216…

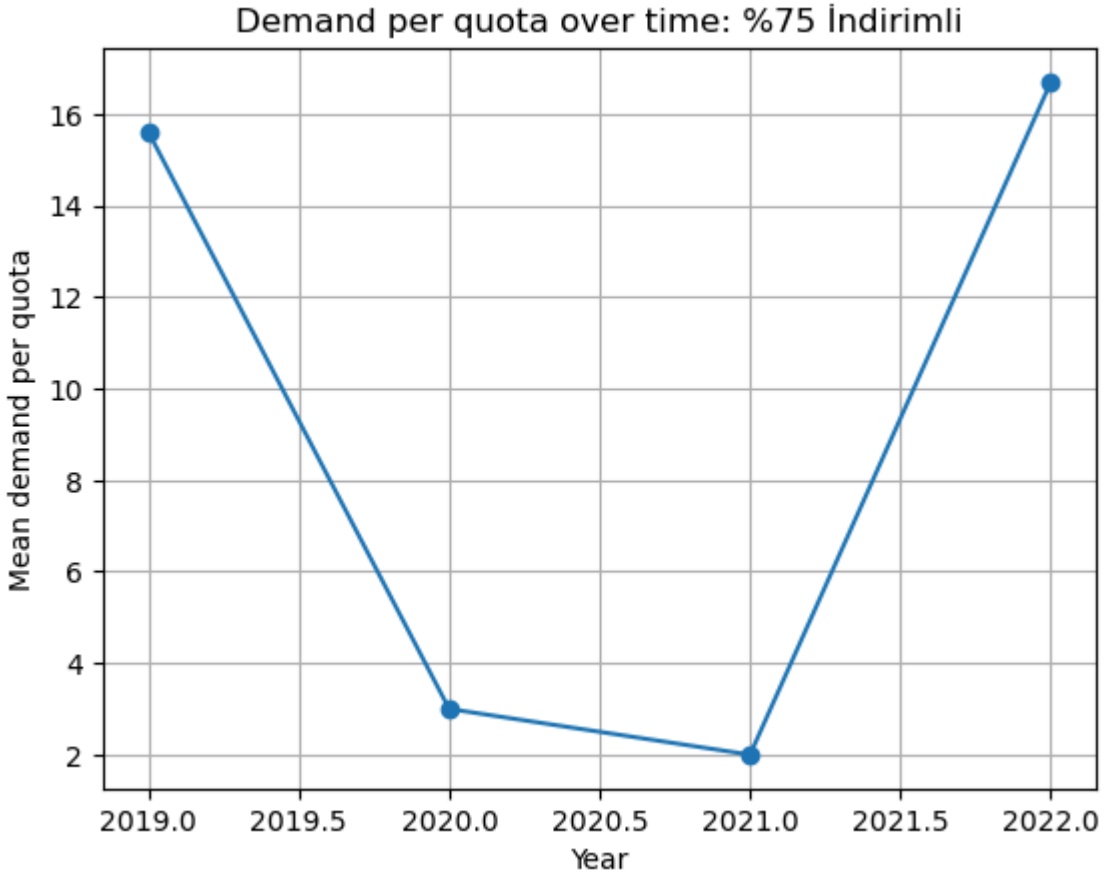
```
demand_by_sch_year = (
    df_core.groupby(["year", "scholarship_type"])["demand_per_quota"]
    .mean()
    .reset_index()
)

main_sch = ["Ücretsiz","Burslu","Ücretli","%50 İndirimli","%75 İndirimli","%25 İndiri
trend = demand_by_sch_year[demand_by_sch_year["scholarship_type"].isin(main_sch)]

for sch in main_sch:
    sub = trend[trend["scholarship_type"] == sch]
    plt.figure()
    plt.plot(sub["year"], sub["demand_per_quota"], marker="o")
    plt.title(f"Demand per quota over time: {sch}")
    plt.xlabel("Year")
    plt.ylabel("Mean demand per quota")
    plt.grid(True)
    plt.show()
```

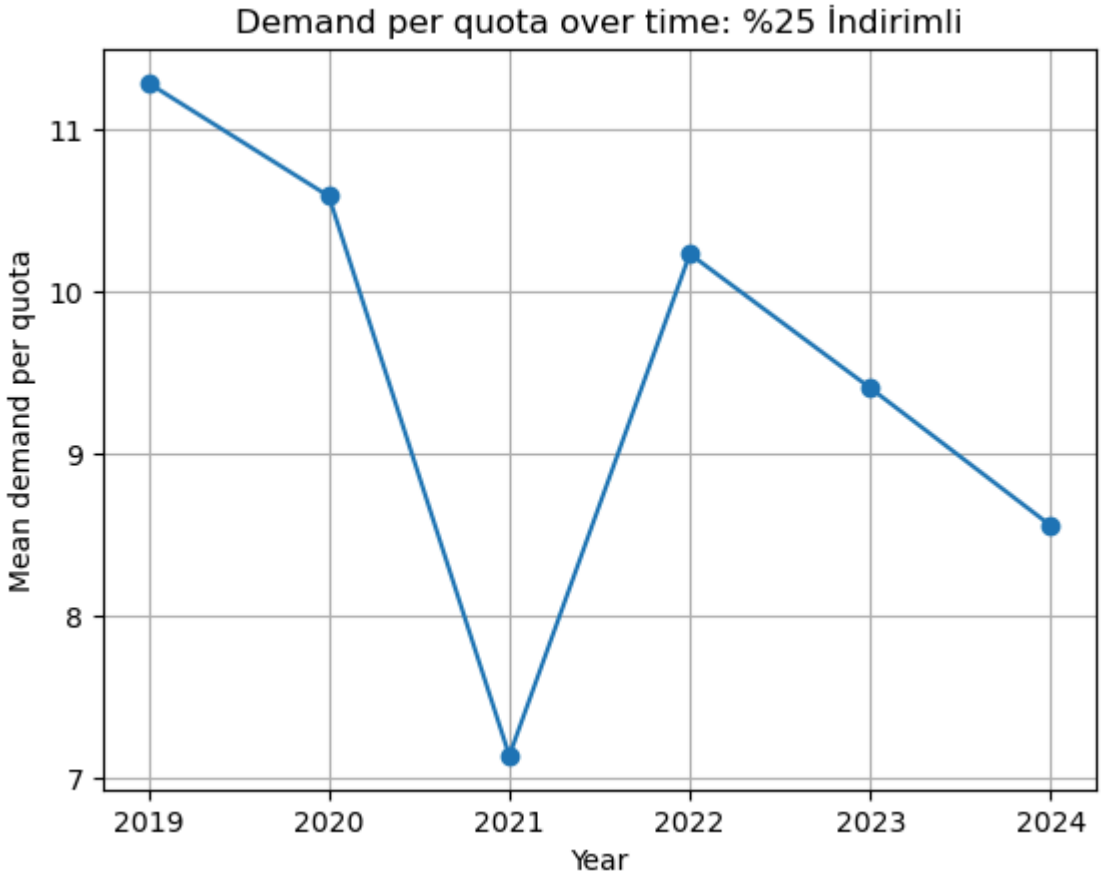



Demand per quota over time: Burslu
Mean demand per quota
38
36
34
32
30
28
26
24
2019
2020
2021
2022
2023
2024
Year


Demand per quota over time: Ücretli
Mean demand per quota
25
20
15
10
5
2019
2020
2021
2022
2023
2024
Year


Demand per quota over time: %50 İndirimli
Mean demand per quota
11
10
9
8
7
6
5
2019
2020
2021
2022
2023
2024
Year


Demand per quota over time: %75 İndirimli
Mean demand per quota
16
14
12
10
8
6
4
2
2019.0
2019.5
2020.0
2020.5
2021.0
2021.5
2022.0
Year



Demand trends over time by scholarship

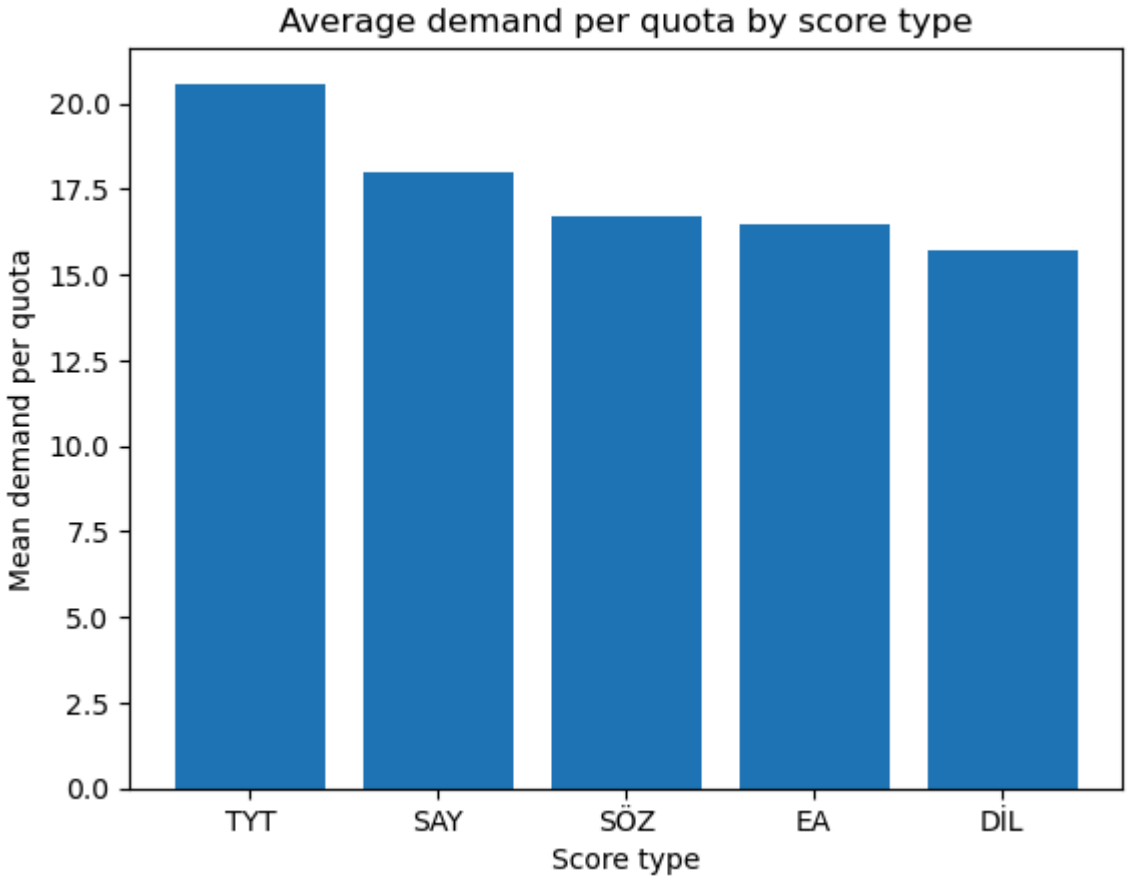
Grouping by scholarship type reveals a clear and intuitive trend: financially favourable programs—particularly Burslu and Ücretsiz—exhibit substantially higher demand per quota. In contrast, partially discounted or fully fee-based programs attract lower interest on average. This gradient suggests that affordability remains a major determinant of student preference behaviour. The differences between the top categories are large enough to indicate true behavioural patterns rather than statistical noise.

Also important that: sample sizes are small quotas vary year-to-year scholarship policies change

In [217…

```
demand_by_score = (
    df_core.groupby("score_type")["demand_per_quota"]
    .agg(["mean","median","std","count"])
    .sort_values("mean", ascending=False)
)
display(demand_by_score)

plt.figure()
plt.bar(demand_by_score.index, demand_by_score["mean"])
plt.title("Average demand per quota by score type")
plt.xlabel("Score type")
plt.ylabel("Mean demand per quota")
plt.show()
```

score_type	mean	median	std	count
TYT	20.583660	15.1	24.217304	60030
SAY	17.986923	13.3	21.820423	29916
SÖZ	16.694727	12.8	16.218540	12288
EA	16.457743	13.1	17.092701	22245
DİL	15.723754	12.5	15.230528	3873



All scholarship types show a clear spike in demand in 2022, indicating a system-wide shift rather than category-specific behaviour.

Burslu consistently has the highest demand across all years, confirming that full scholarships strongly attract applicants.

Ücretsiz programs also maintain high and stable demand, reinforcing the importance of affordability.

Ücretli and partial-discount categories remain much less competitive, with flatter trends and more variability—evidence of strong price sensitivity.

Small categories like %25 and %75 İndirimli show irregular patterns due to limited sample sizes.

Overall, financial accessibility is the primary driver of multi-year demand trends, overshadowing year-to-year fluctuations and score-type effects.

In [218…

```
interaction = (
    df_core.groupby(["scholarship_type","score_type"])["demand_per_quota"]
    .mean()
    .reset_index()
    .pivot(index="scholarship_type", columns="score_type", values="demand_per_quota")
)
```

```
display(interaction)
```

score_type scholarship_type	DİL	EA	SAY	SÖZ	TYT
%25 İndirimli	4.568116	10.727009	9.580299	4.992466	10.200203
%50 İndirimli	5.870732	7.860192	8.671183	5.169680	8.352388
%75 İndirimli	17.759259	10.477916	13.811839	11.196891	19.019604
AÖ-Ücretli	NaN	20.102165	42.500000	12.872549	40.057853
Burslu	26.931392	25.962723	28.749909	25.190350	33.692456
UE-Ücretli	NaN	3.125000	1.033333	NaN	NaN
UÖ-Ücretli	7.700000	17.052632	10.925000	14.487500	27.043636
Ücretli	7.790496	12.516007	11.835529	8.378660	16.423454
Ücretsiz	16.640770	17.037615	18.307814	19.442045	20.898247
İÖ-Ücretli	12.097857	11.670324	11.258091	13.043760	17.282766

Score type effects and interaction with scholarship

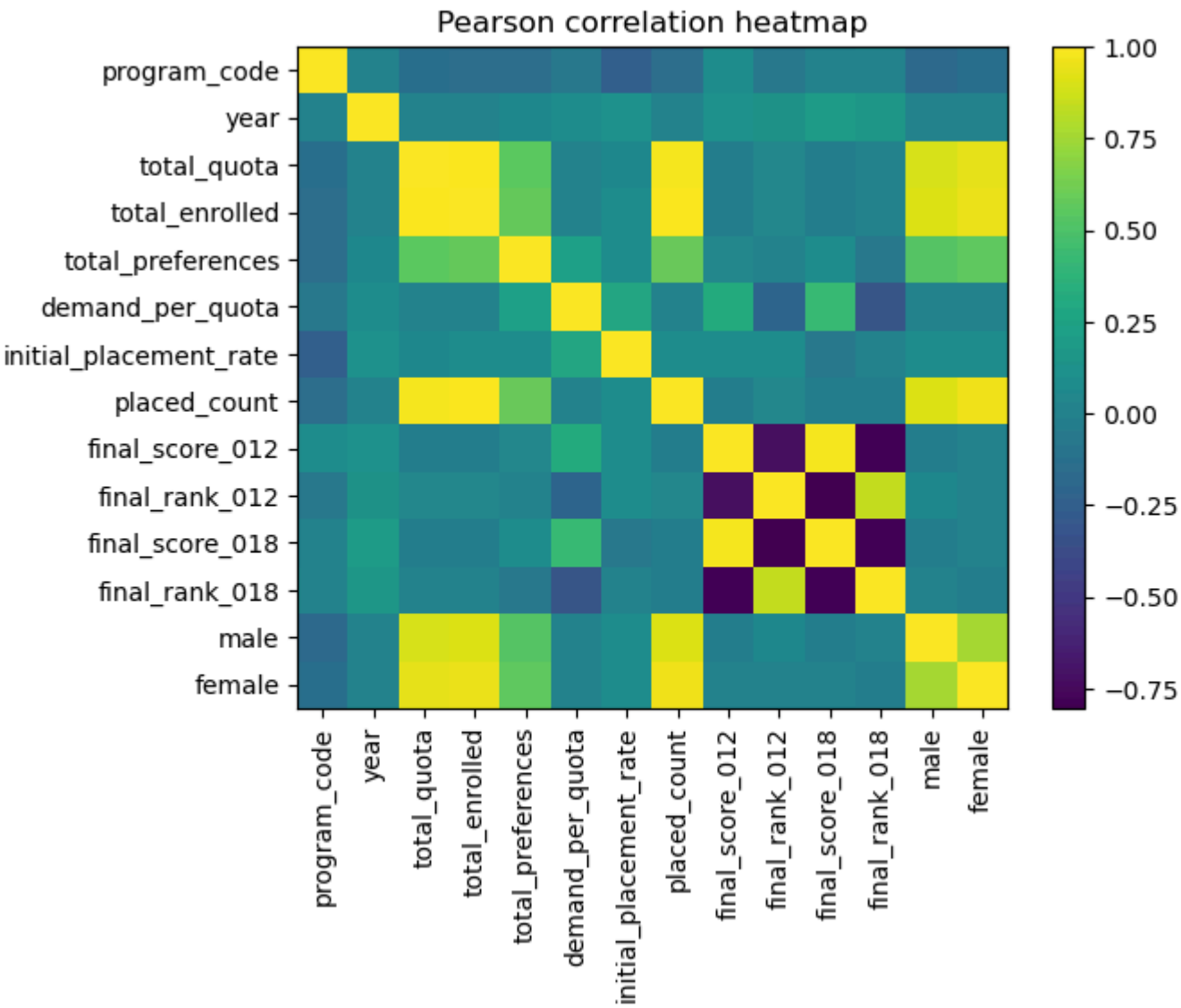
Differences across score types exist but are less pronounced than across scholarship categories. TYT programs, being the broadest and most accessible entry route, tend to show slightly higher demand levels. In contrast, specialised score types such as DİL or SAY display lower or more variable demand, likely reflecting differences in programme diversity and national labour-market demand. However, the effect sizes are modest relative to scholarship-based differences, suggesting that financial accessibility exerts a stronger influence on student choices than exam route.

In [219…

```
corr_pearson = num_df.corr(method="pearson")
corr_spearman = num_df.corr(method="spearman")

display(corr_pearson)
display(corr_spearman)

plt.figure()
plt.imshow(corr_pearson, aspect="auto")
plt.colorbar()
plt.title("Pearson correlation heatmap")
plt.xticks(range(len(corr_pearson.columns)), corr_pearson.columns, rotation=90)
plt.yticks(range(len(corr_pearson.columns)), corr_pearson.columns)
plt.show()
```

	program_code	year	total_quota	total_enrolled	total_preferences	de
program_code	1.000000	-0.001860	-0.140013	-0.145934	-0.153990	
year	-0.001860	1.000000	0.006184	0.014743	0.054445	
total_quota	-0.140013	0.006184	1.000000	0.988859	0.555717	
total_enrolled	-0.145934	0.014743	0.988859	1.000000	0.580368	
total_preferences	-0.153990	0.054445	0.555717	0.580368	1.000000	
demand_per_quota	-0.065537	0.085828	-0.004417	0.009095	0.220634	
initial_placement_rate	-0.228252	0.110274	0.051114	0.094563	0.088070	
placed_count	-0.145760	0.015052	0.975492	0.985998	0.588050	
final_score_012	0.070670	0.116862	-0.023089	-0.015769	0.044587	
final_rank_012	-0.082069	0.121582	0.033399	0.036679	-0.001191	
final_score_018	0.007726	0.209082	-0.017669	-0.017551	0.072304	
final_rank_018	0.027286	0.176753	-0.005827	-0.005878	-0.073667	
male	-0.161601	0.007049	0.891553	0.901748	0.534063	
female	-0.120471	0.019004	0.935608	0.945281	0.566314	

	program_code	year	total_quota	total_enrolled	total_preferences	de
program_code	1.000000	0.017811	-0.549776	-0.532514	-0.532275	
year	0.017811	1.000000	0.039660	0.071166	0.152238	
total_quota	-0.549776	0.039660	1.000000	0.950079	0.743822	
total_enrolled	-0.532514	0.071166	0.950079	1.000000	0.819615	
total_preferences	-0.532275	0.152238	0.743822	0.819615	1.000000	
demand_per_quota	-0.127346	0.172157	0.019696	0.161568	0.620832	
initial_placement_rate	-0.173780	-0.020875	0.242349	0.445344	0.479025	
placed_count	-0.525324	0.075899	0.936575	0.985719	0.835993	
final_score_012	0.094826	0.143803	-0.050775	-0.016166	0.232156	
final_rank_012	-0.104447	0.090145	0.070072	0.078318	-0.064657	
final_score_018	0.034364	0.243507	0.057191	0.057877	0.393426	
final_rank_018	-0.028251	0.159583	-0.059561	-0.060302	-0.262500	
male	-0.512130	0.035568	0.785318	0.828001	0.682825	
female	-0.418127	0.095830	0.802637	0.853713	0.753790	



Most variables show weak to moderate correlations, indicating that key admissions metrics (quota, preferences, ranks, gender counts) capture distinct aspects of program competitiveness.

Demand_per_quota correlates positively with total_preferences and negatively with final_rank, meaning higher-demand programs tend to attract more applicants and have stronger (lower-numbered) entrance ranks.

Quota is weakly related to demand, showing that program size does not strongly determine competitiveness.

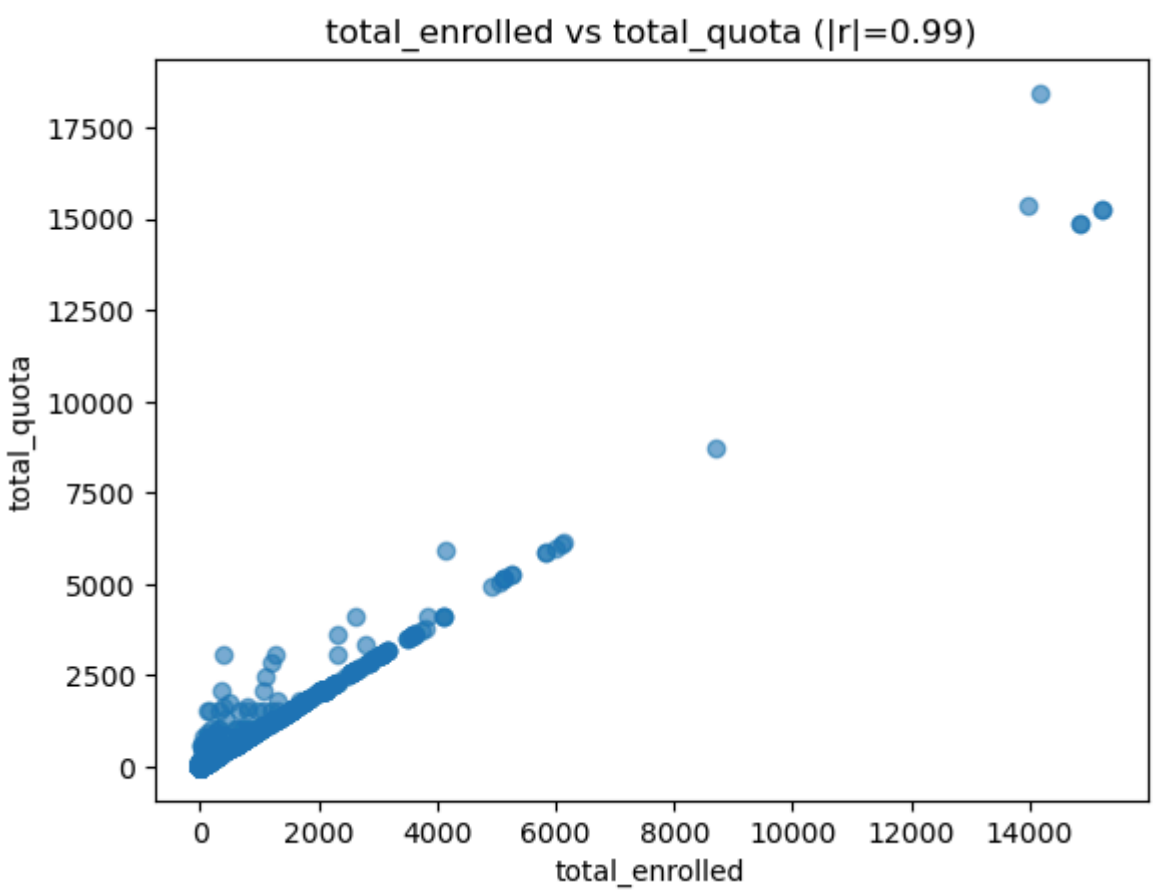
Spearman correlations are slightly stronger than Pearson, reflecting the non-linear and heavy-tailed nature of many variables.

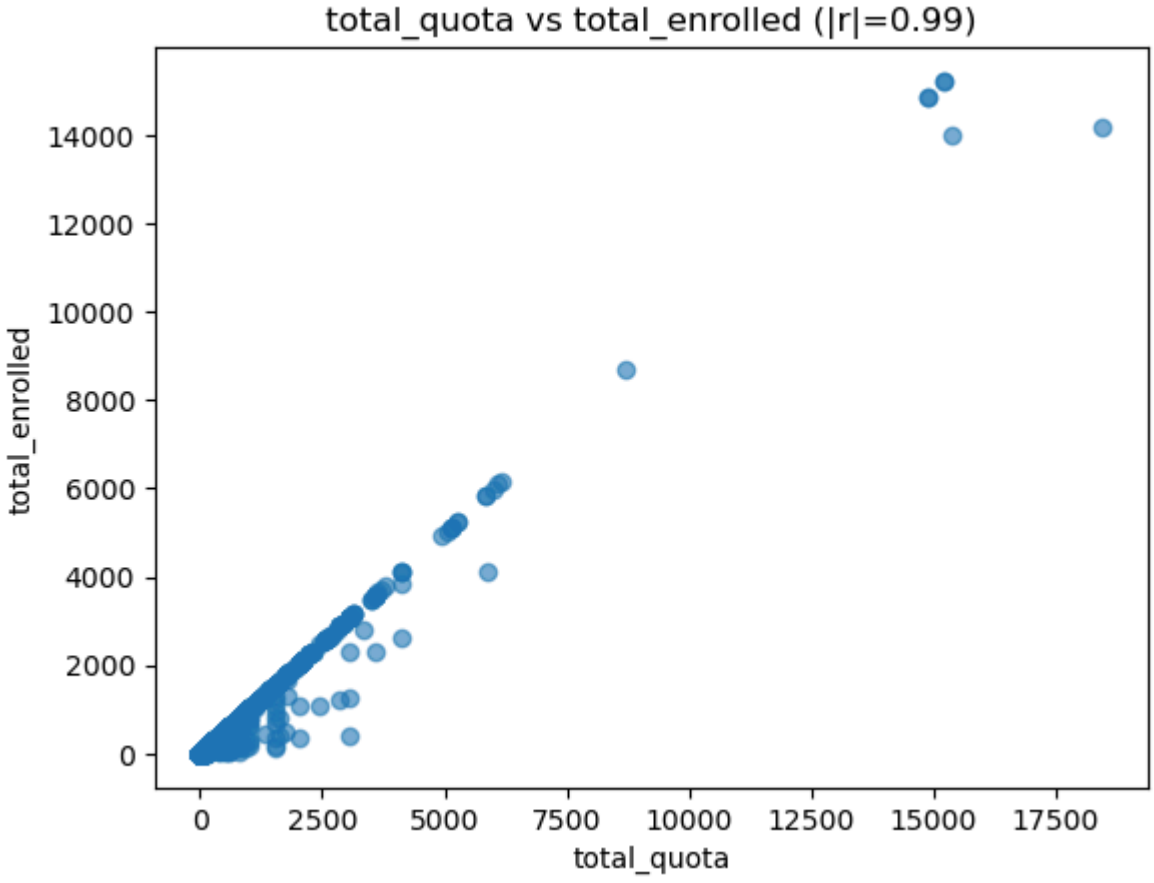
Overall, the correlation structure suggests a complex admissions landscape where no single numeric factor fully explains program demand.

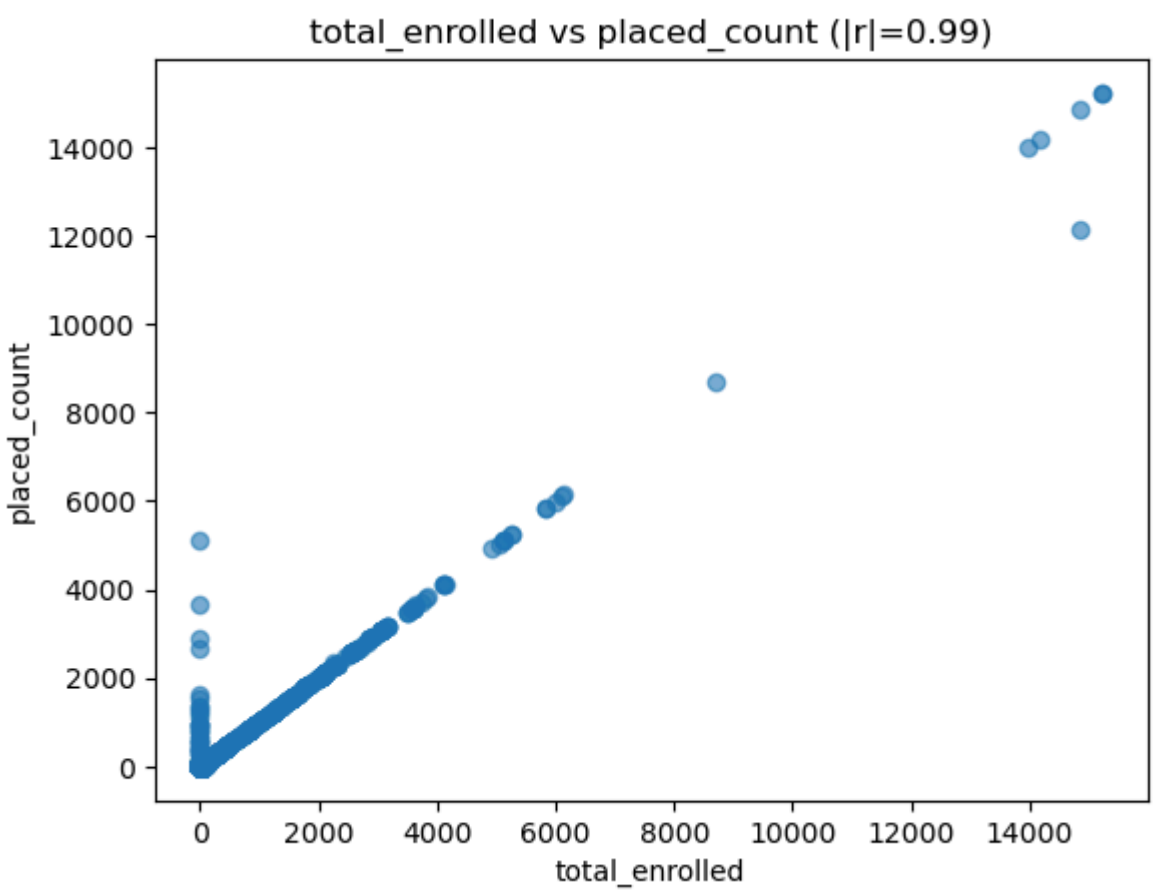
In [ ]:

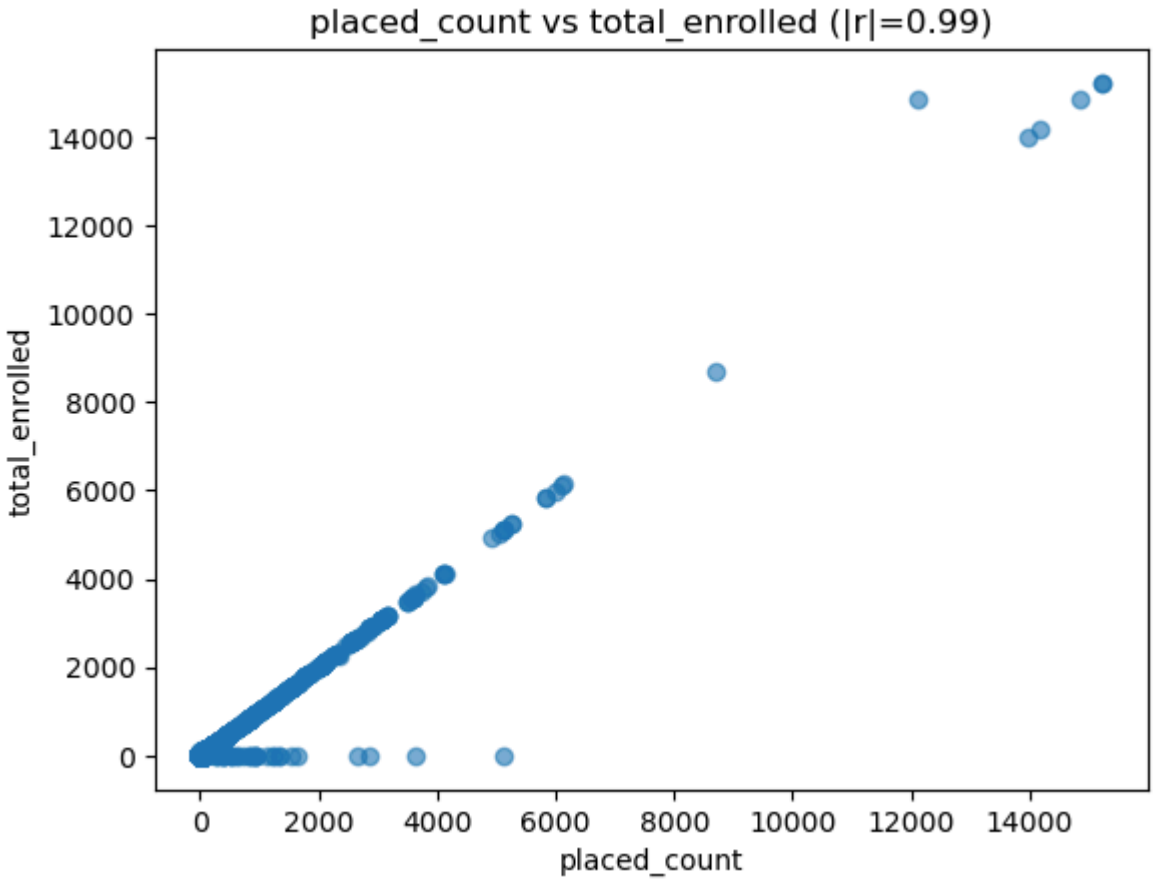
In [220…

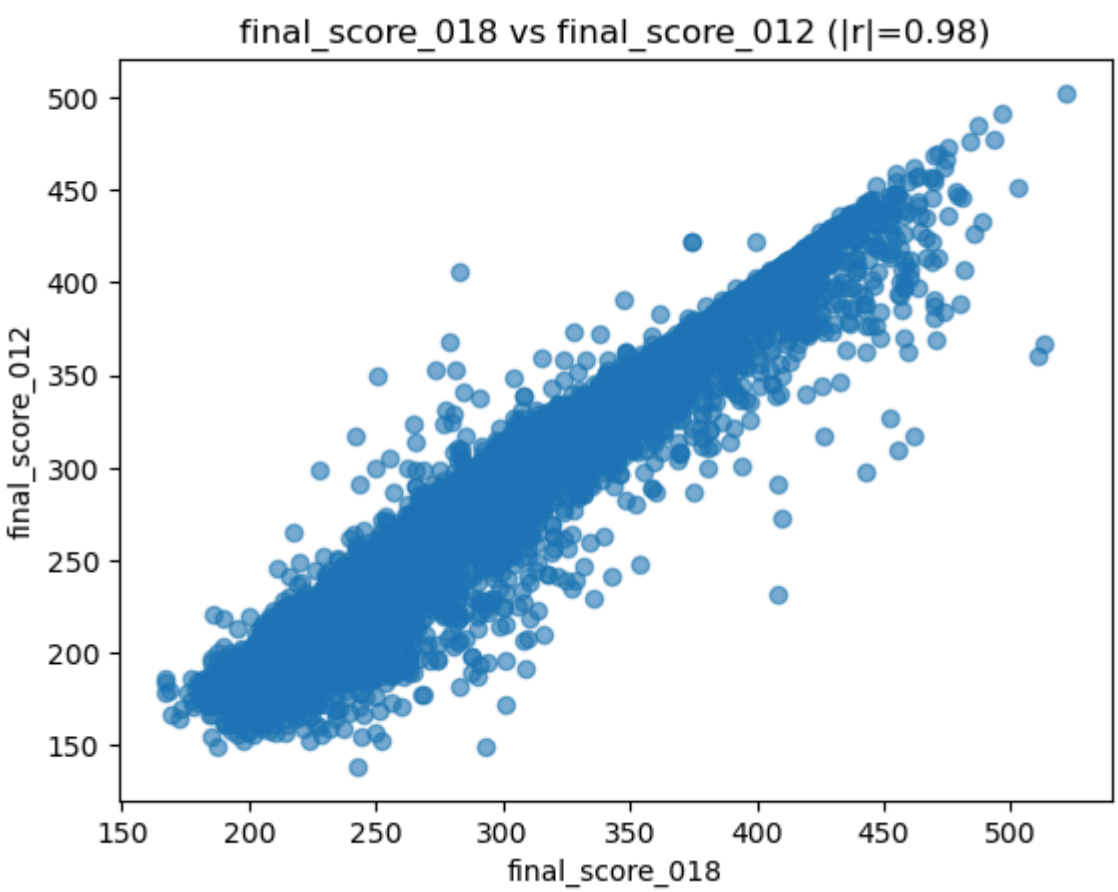
```python
for (x, y), val in top_pairs.items():
    plt.figure()
    plt.scatter(df_core[x], df_core[y], alpha=0.6)
    plt.title(f"{x} vs {y} (|r|={val:.2f})")
    plt.xlabel(x)
    plt.ylabel(y)
    plt.show()
```


total_enrolled vs total_quota (|r|=0.99)
total_quota
total_enrolled
0
2500
5000
7500
10000
12500
15000
17500
0
2000
4000
6000
8000
10000
12000
14000


total_quota vs total_enrolled (|r|=0.99)
total_enrolled
total_quota
0
2000
4000
6000
8000
10000
12000
14000
0
2500
5000
7500
10000
12500
15000
17500


total_enrolled vs placed_count (|r|=0.99)
placed_count
total_enrolled
0
2000
4000
6000
8000
10000
12000
14000


placed_count vs total_enrolled (|r|=0.99)
total_enrolled
placed_count
0
2000
4000
6000
8000
10000
12000
14000



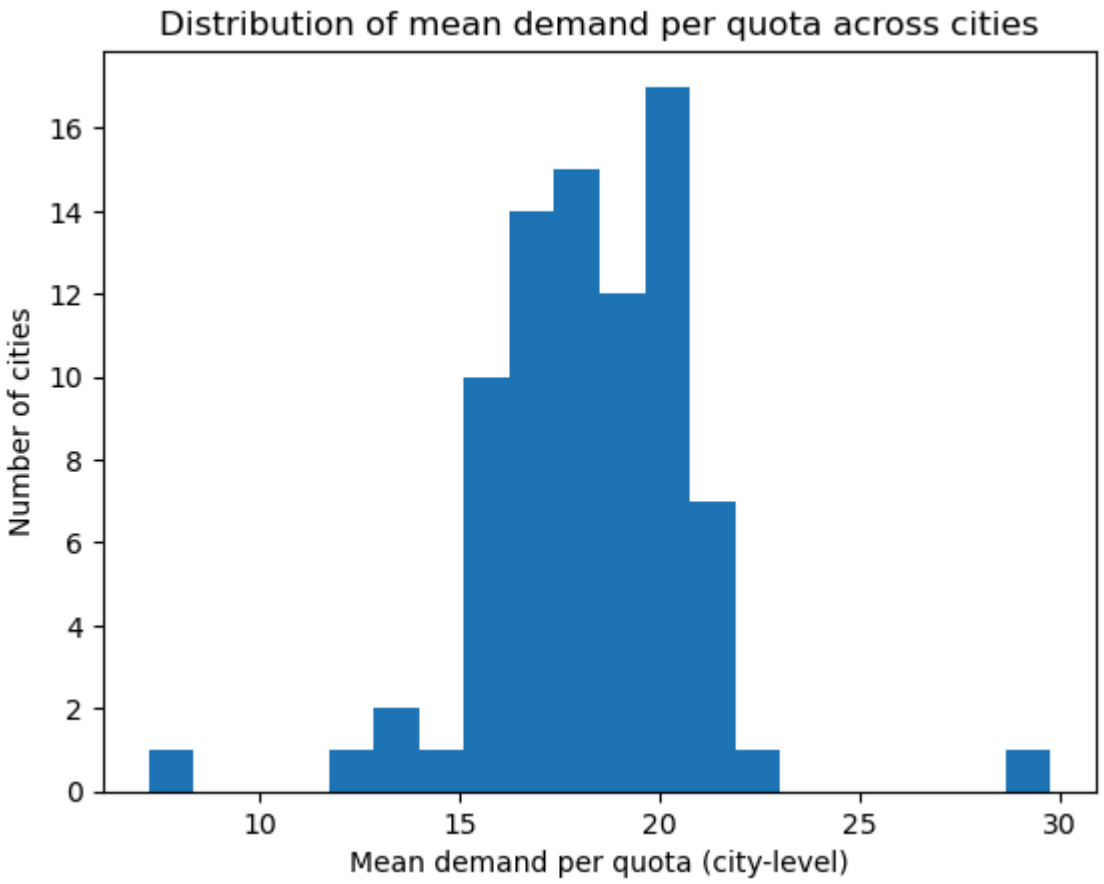
Scatterplots of the strongest correlated variable pairs show clear and interpretable patterns. Demand-related variables (such as total preferences and demand per quota) move together positively, indicating that applicant interest and competitiveness reinforce each other. Rank-based variables show strong negative relationships with demand, reflecting the link between program popularity and entrance exam competitiveness.

Some correlations arise from structural relationships, such as quota and total enrolment.

In [221…

```
city_demand = (
    df_core.groupby("city")["demand_per_quota"]
    .mean()
    .sort_values(ascending=False)
)

display(city_demand.head(15))
display(city_demand.describe())

plt.figure()
plt.hist(city_demand, bins=20)
plt.title("Distribution of mean demand per quota across cities")
plt.xlabel("Mean demand per quota (city-level)")
plt.ylabel("Number of cities")
plt.show()
```

```
city
ESKİŞEHİR    29.776429
BURSA        22.052569
ANKARA       21.836819
AMASYA       21.713428
MUĞLA        21.444404
BAYBURT      21.365060
BALIKESİR    21.307849
HAKKARİ      21.098058
YOZGAT       20.987163
İSTANBUL     20.588939
SİNOP        20.492906
BOLU         20.487055
SAMSUN       20.429070
ANTALYA      20.370963
ARTVİN       20.210973
Name: demand_per_quota, dtype: float64
count    82.000000
mean     18.192723
std       2.710300
min       7.200472
25%      16.578307
50%      18.182964
75%      19.808068
max      29.776429
Name: demand_per_quota, dtype: float64
```



The cities with the highest average demand per quota (top 15 displayed) tend to be those hosting large, reputable, or highly competitive universities. This indicates that institutional prestige and program mix, rather than geography alone, largely drive city-level demand.

The descriptive statistics show a moderate spread across cities: while some cities consistently attract much higher demand, most sit near the overall mean. This suggests regional variation exists, but it is not extreme.

The histogram reveals that the distribution of mean demand across cities is moderately right-skewed: most cities cluster around mid-range demand levels, with only a small number standing out

as high-demand hubs.

City-level means are influenced by how many programs each region offers. Overall, city-level demand appears to reflect concentration of high-demand programs, not inherent regional differences, pointing toward institutional characteristics as the primary factor shaping competitiveness rather than location alone.

Limitations

This is an exploratory analysis; all relationships are descriptive and patterns between demand, scholarship, score type, or rank cannot be interpreted causally.

Aggregation at the program-year level may obscure department-level nuances (e.g., sub-department variations, evening vs daytime programs, program-specific specialization tracks).

Scholarship trends can reflect institutional strategies, not only student behaviour (e.g., universities adjusting scholarship availability in certain years).

City-level comparisons reflect composition effects (presence of large universities or prestigious faculties), not necessarily inherent regional demand differences.

As well as those mentioned at the start of the notebook relating more closely to data used.

Conclusion

The analysis shows that scholarship type is the strongest predictor of demand per quota, with fully funded and free public programs (Burslu and Ücretsiz) consistently attracting the highest levels of student demand. This pattern persists across all score types, indicating that financial accessibility remains a structural driver of student preference behaviour rather than a domain-specific effect.

Score type differences exist but are relatively modest, with TYT programs showing slightly higher demand on average due to their broad applicability and wider range of participating programs. However, the magnitude of these differences is small compared with scholarship effects, suggesting that affordability influences admissions competitiveness more strongly than exam pathway selection.

A notable system-wide shift emerges in 2022, where demand per quota sharply increases across many categories before stabilising at a higher level in 2023–2024. This pattern indicates a macro-level change in the admissions environment—potentially linked to demographic shifts, policy changes, or economic conditions—rather than variation attributable to particular program types or institutions.

City-level patterns show meaningful but interpretable variation: some urban centres consistently exhibit higher demand, likely reflecting the concentration of prestigious institutions rather than purely regional preference. This underscores that regional inequalities in admissions competitiveness are often compositional, shaped by the distribution of high-demand universities rather than underlying geographic effects.

Together, these findings highlight that scholarship structures, national-level shifts in demand, and institutional clustering shape student decision-making more strongly than exam score type alone. Understanding these dynamics is crucial for policymakers and institutions aiming to balance financial accessibility, program capacity, and regional educational equity.

Further data and analysis needed

Access to program-level reputation indicators, employment outcomes, or institutional rankings would allow clearer interpretation of why certain programs attract disproportionate demand.

Individual-level preference data or socioeconomic indicators would help assess whether scholarship-driven demand reflects financial need, perceived value, or broader market conditions.

A causal understanding would benefit from year-specific policy documentation, demographic data (e.g., applicant pool size), and time-series modelling to examine structural breaks more formally.