**17MDC55 – HUMAN RESOURCE DEVELOPMENT LAB**

**CAREER ANALYTICS**

Campus recruitment is a strategy for sourcing, engaging and hiring young talent for internship and entry-level positions. College recruiting is typically a tactic for medium- to large-sized companies with high-volume recruiting needs, but can range from small efforts (like working with university career centers to source potential candidates) to large-scale operations (like visiting a wide array of colleges and attending recruiting events throughout the spring and fall semester). Campus recruitment often involves working with university career services centres and attending career fairs to meet in-person with college students and recent graduates.

**Context of our Dataset:**

Our dataset revolves around the placement season of a Business School in India. Where it has various factors on candidates getting hired such as work experience, exam percentage etc., Finally it contains the status of recruitment and remuneration details.

**Goals**

There are three primary goals of this kernel. Doing a exploratory analysis and visualization analysis of the Recruitment dataset

**Prediction:**

To predict whether a student got placed or not using classification models

**1.IMPORTING LIBRARIES:**

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

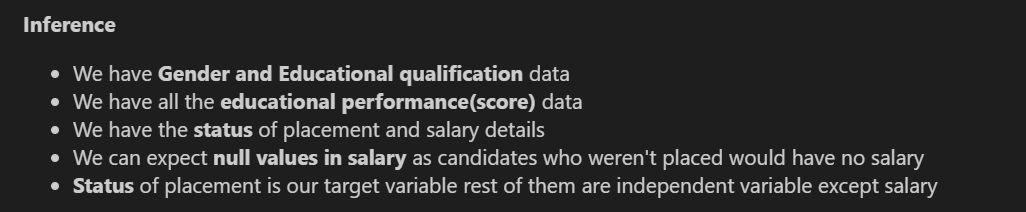
import seaborn as sns

import sklearn

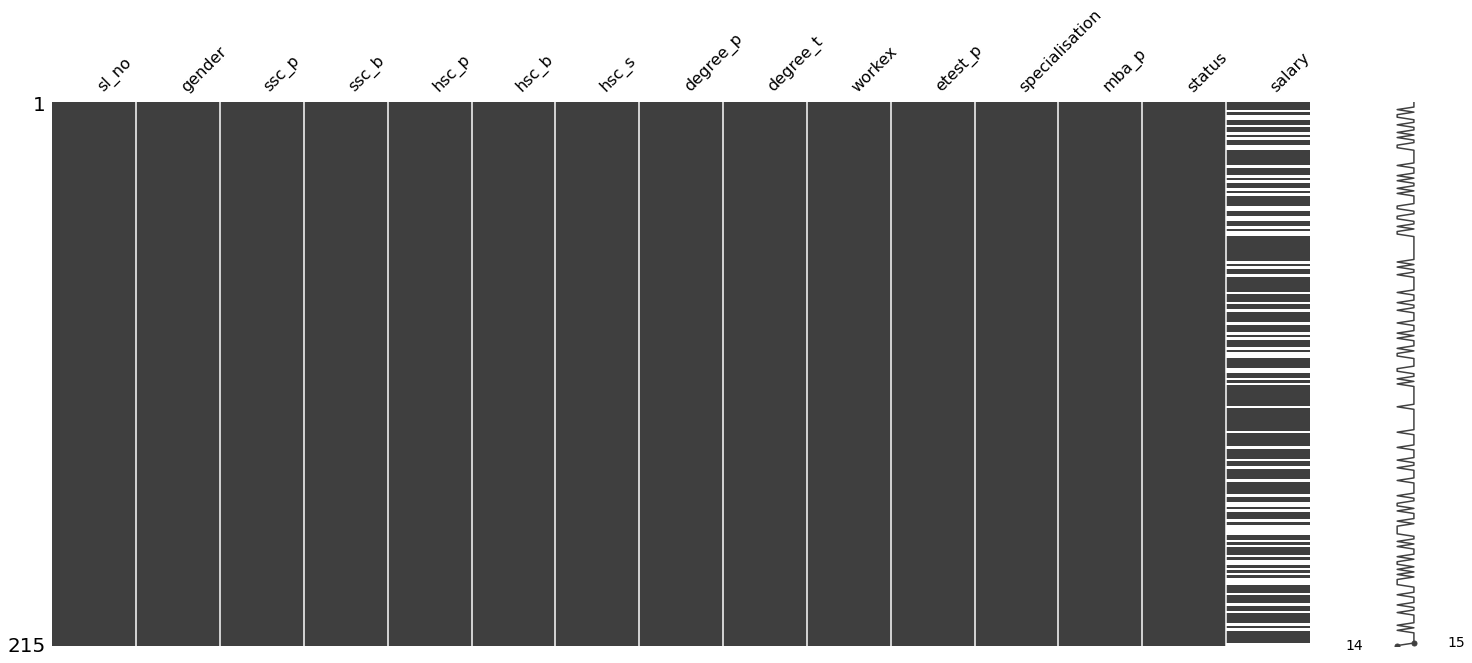
%matplotlib inline

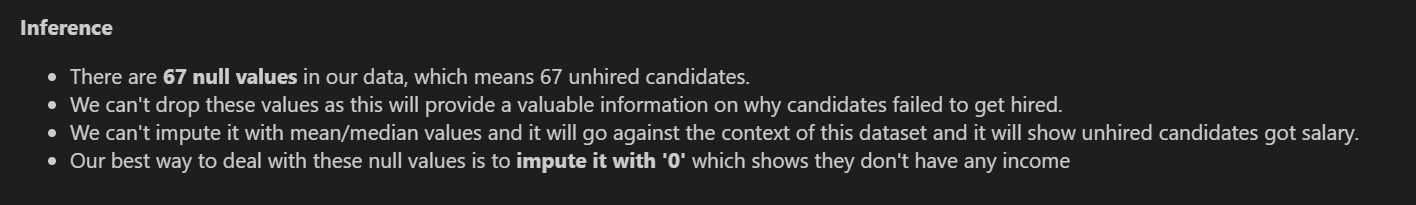
**2.EXTRACTING AND EXAMINING DATASET:**





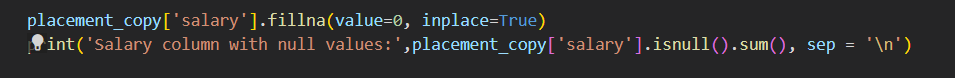
**3.CHECKING MISSING DATA:**



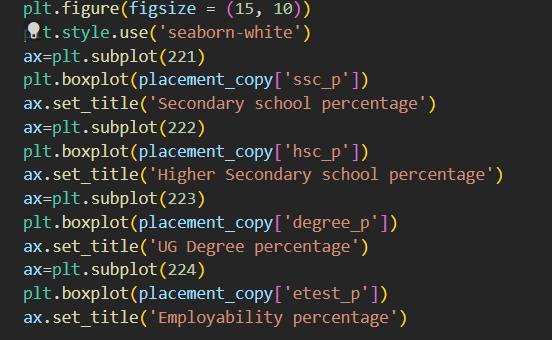


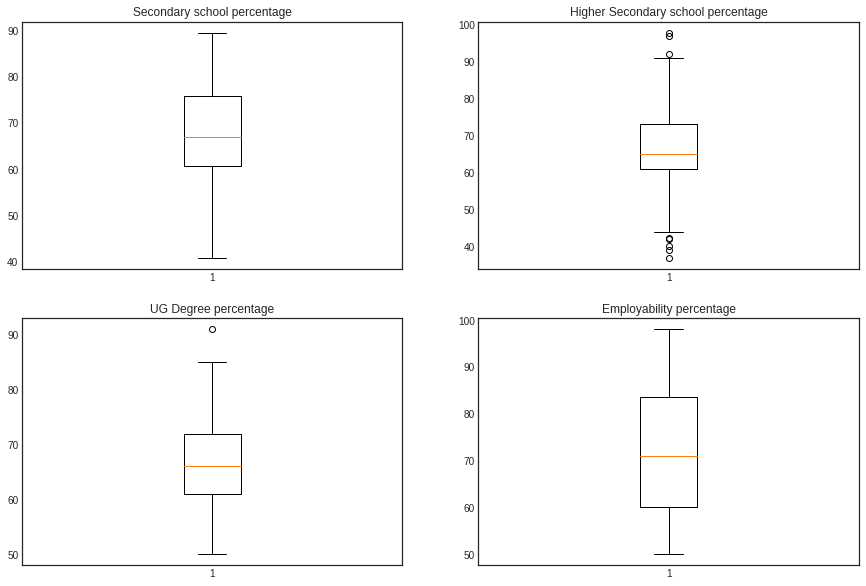
**4.DATA CLEANING**

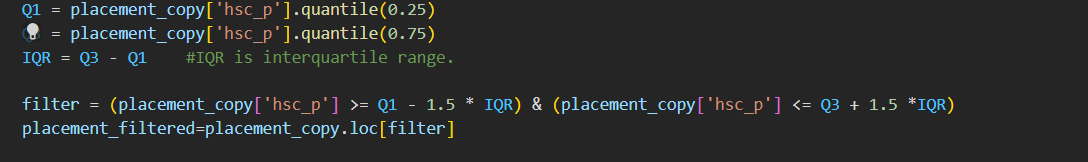
HANDLING MISSING VALUES:

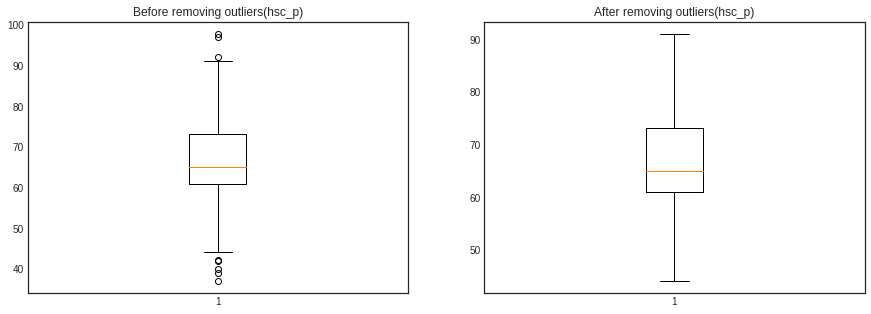


**5.OUTLIERS:**





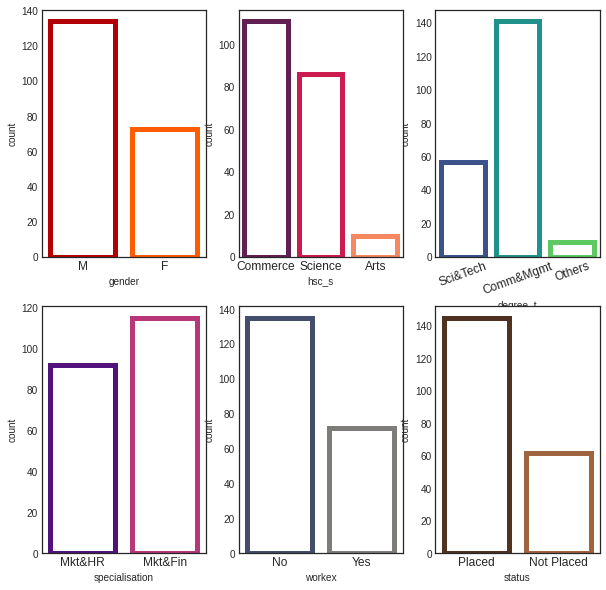


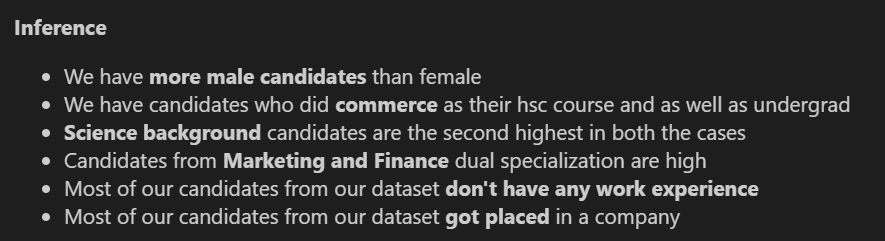


**6.DATA VISUALIZATION:**

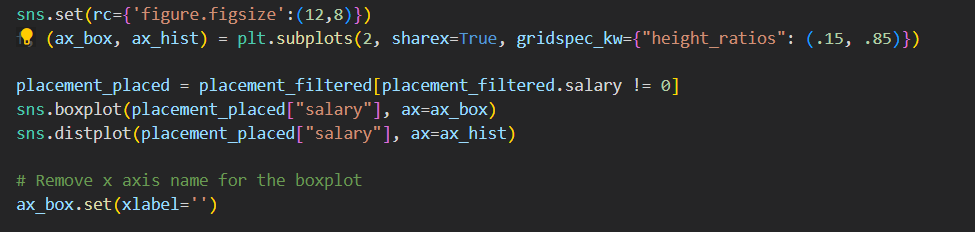
COUNT OF CATEGORICAL FEATURES – COUNT PLOT

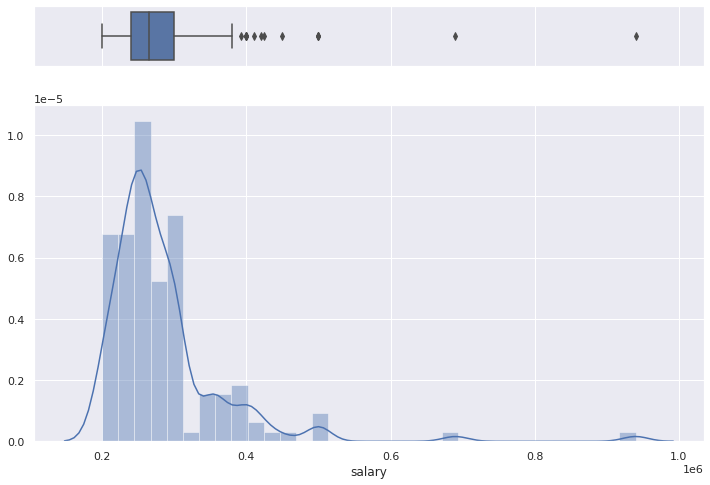


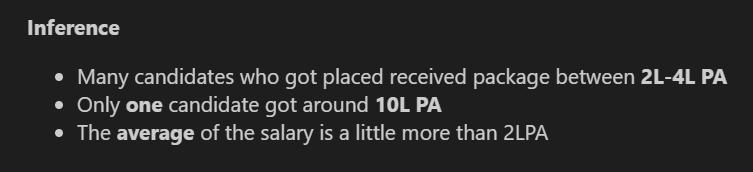




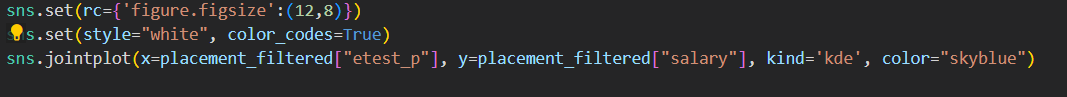
DISTRIBUTION SALARY – PLACED STUDENTS:

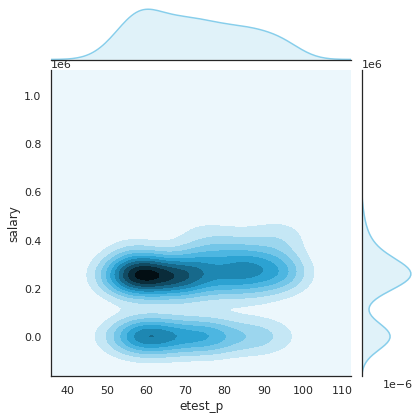


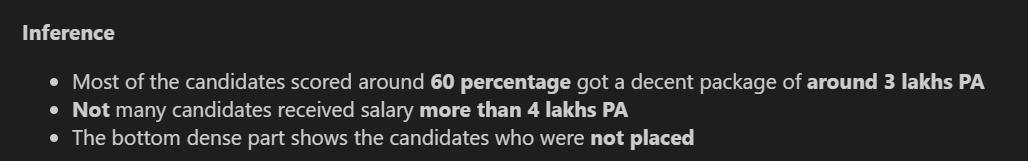




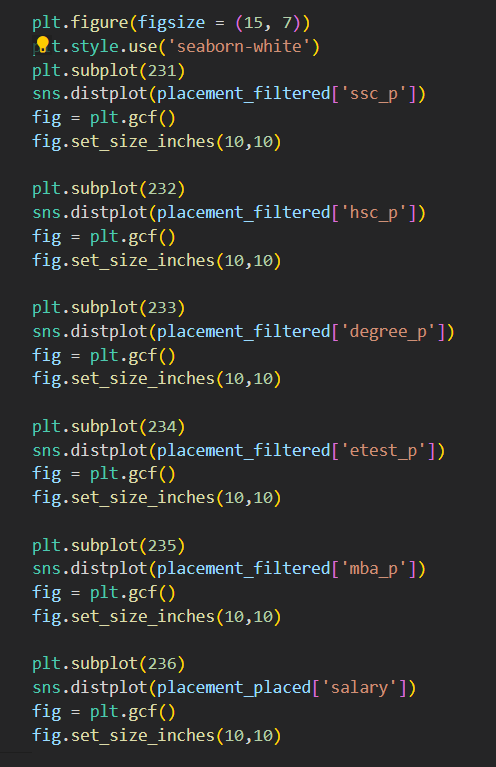
EMPLOYABILITY SCORE AND SALARY-JOINT PLOT:

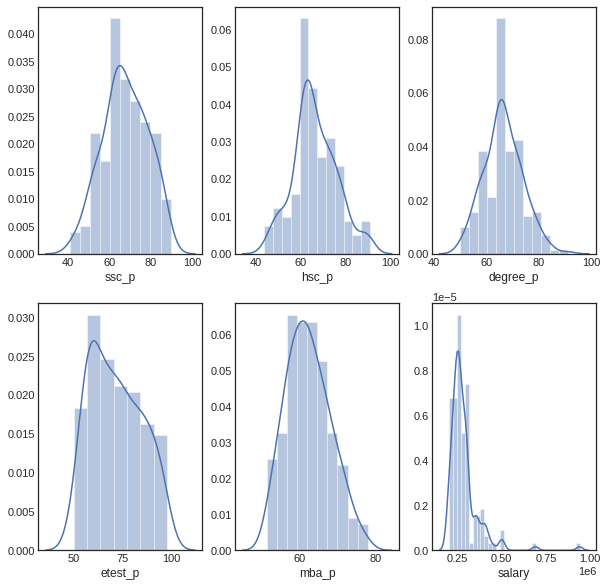


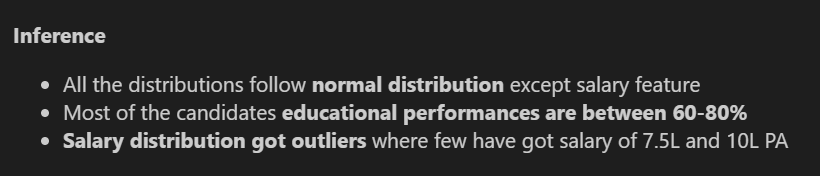




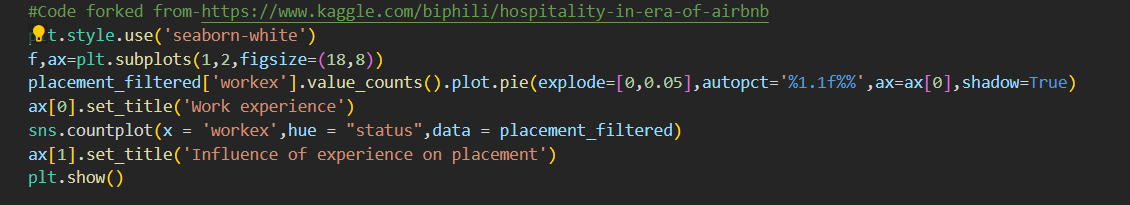
DISTRIBUTION OF ALL PERCENTAGES:

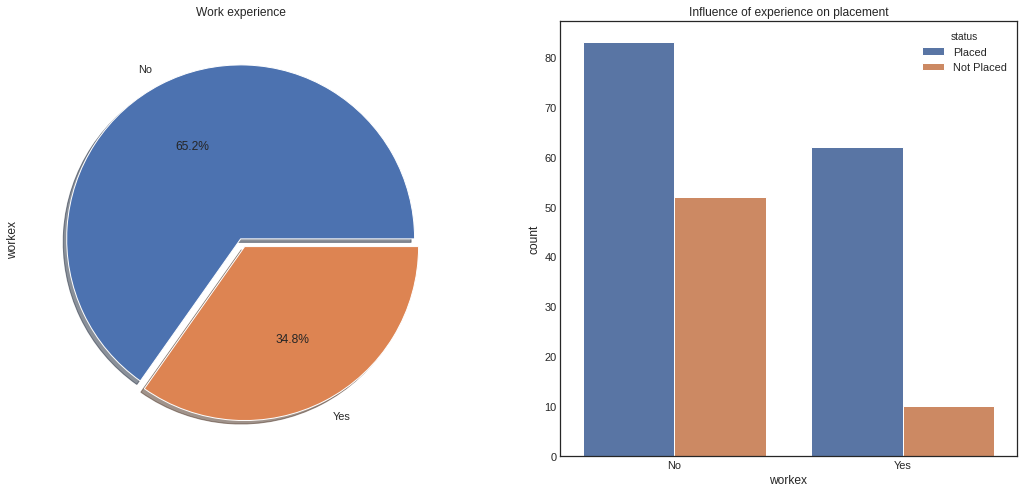
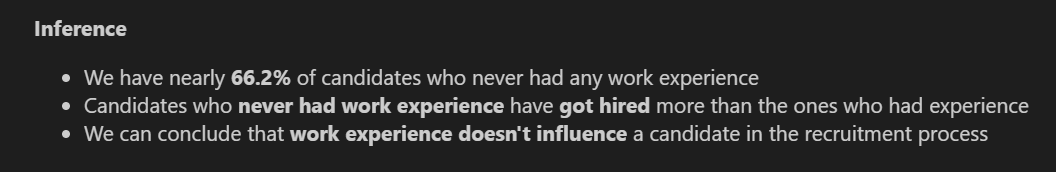




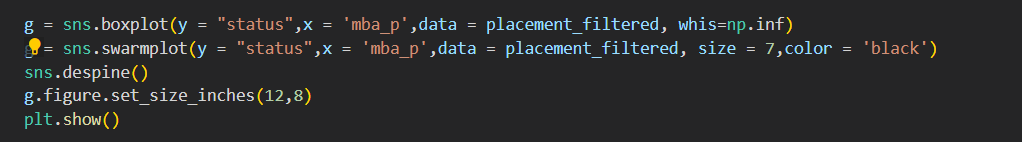


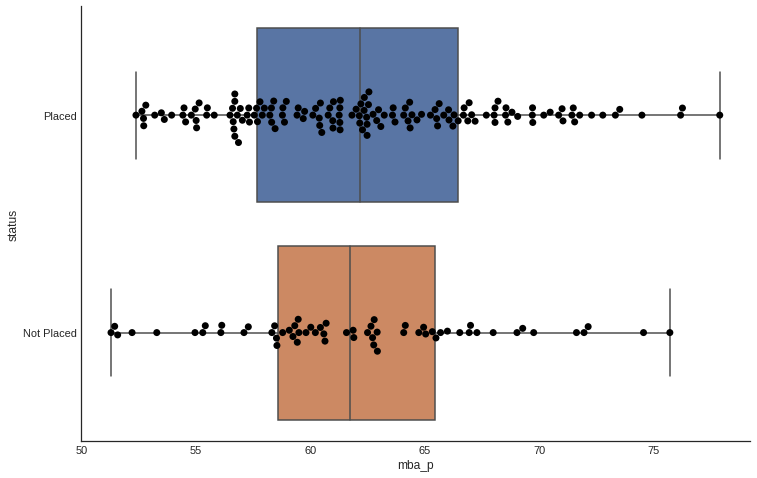
WORK EXPERIENCE VS PLACEMENT STATUS:

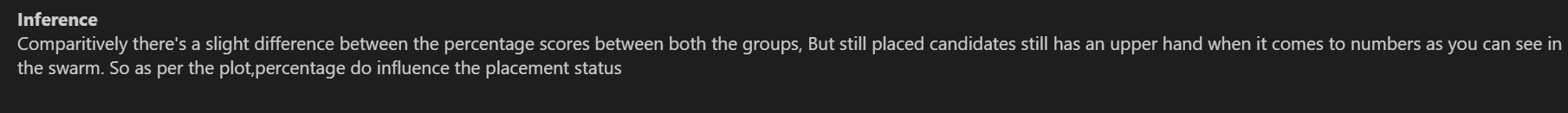


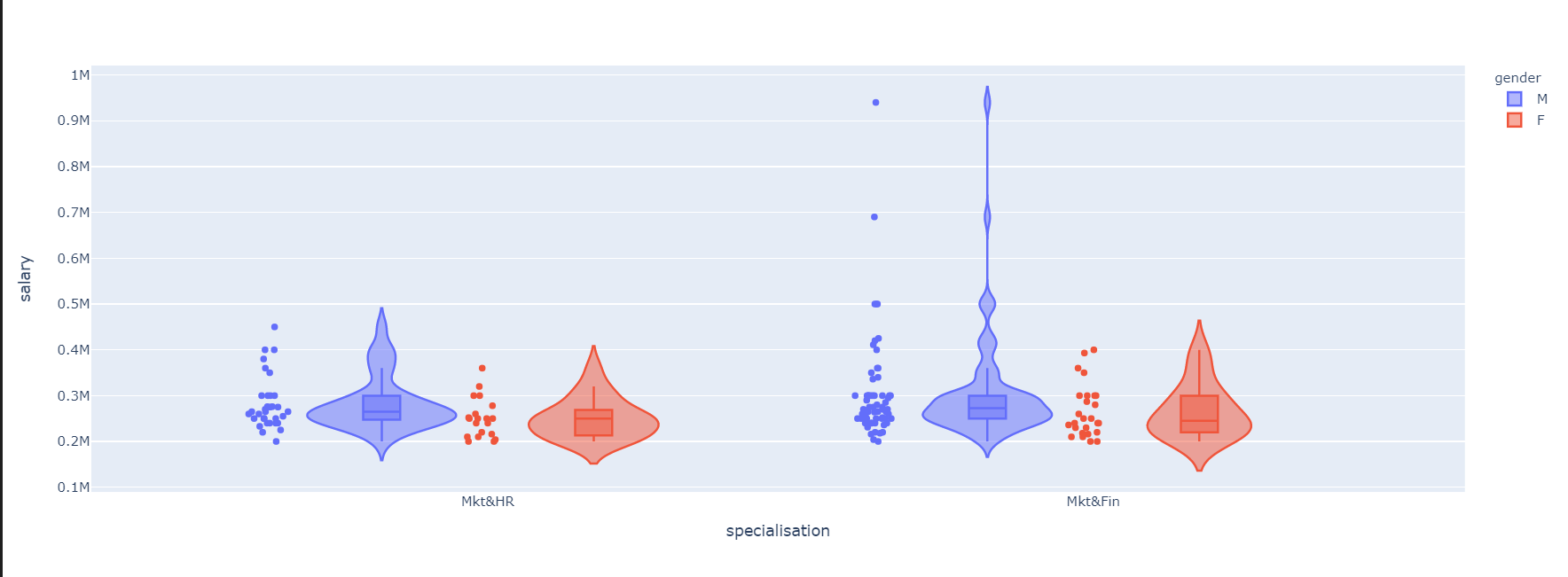
MBA MARK VS PLACEMENT STATUS – DOES YOUR ACADEMIC SCORE INFLUENCE:

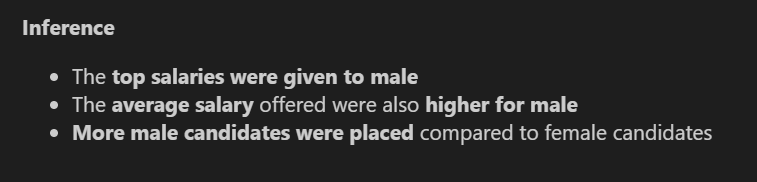






IS THERE ANY GENDER BIAS WHILE OFFERING RENUMERATION:





**MACHINE LEARNING MODELS:**

**LOGISTIC REGRESSION:**

from sklearn.linear\_model import LogisticRegression

from sklearn import metrics

logreg = LogisticRegression()

logreg.fit(X\_train, y\_train)

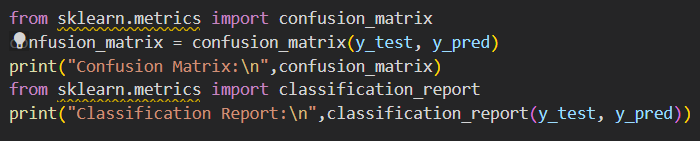
y\_pred = logreg.predict(X\_test)

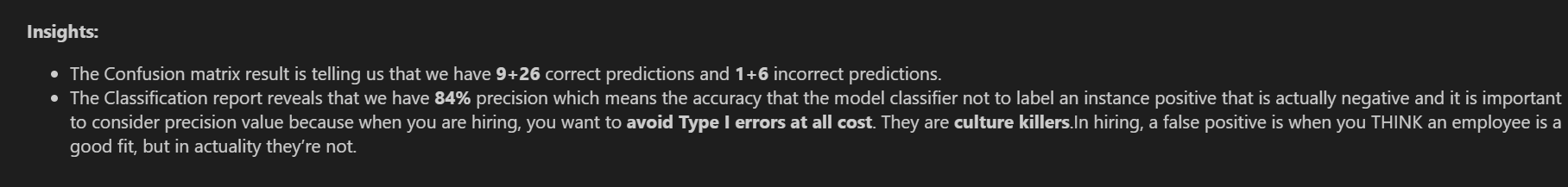
print('Accuracy of logistic regression classifier on test set: {:.2f}'.format(logreg.score(X\_test, y\_test)))

Accuracy of logistic regression classifier on test set: 0.81

83% accurate

Confusion matrix and see the classification report





**DECISION TREE:**

from sklearn.tree import DecisionTreeClassifier

dt = DecisionTreeClassifier(criterion="gini", max\_depth=3)

dt = dt.fit(X\_train,y\_train)

y\_pred = dt.predict(X\_test)

print("Accuracy:",metrics.accuracy\_score(y\_test, y\_pred))

from sklearn.externals.six import StringIO

from sklearn.tree import export\_graphviz

from IPython.display import Image

import pydotplus

dot\_data = StringIO()

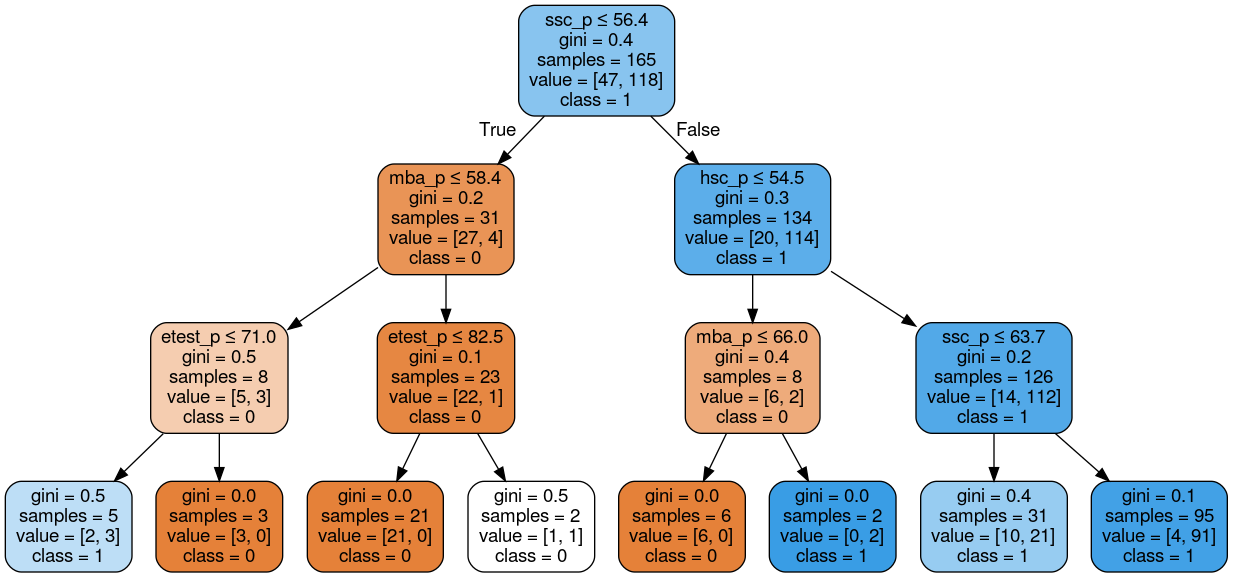
export\_graphviz(dt, out\_file=dot\_data,

                filled=True, rounded=True,

                special\_characters=True,feature\_names = feature\_cols,class\_names=['0','1'], precision=1)

graph = pydotplus.graph\_from\_dot\_data(dot\_data.getvalue())

Image(graph.create\_png())



**RANDOM FOREST**

from sklearn.ensemble import RandomForestClassifier

rt=RandomForestClassifier(n\_estimators=100)

rt.fit(X\_train,y\_train)

y\_pred=rt.predict(X\_test)

from sklearn import metrics

print("Accuracy:",metrics.accuracy\_score(y\_test, y\_pred))

Accuracy: 0.8333333333333334

**K NEAREST NEIGHBOURS:**

from sklearn.neighbors import KNeighborsClassifier

error\_rate = []

for i in range(1,40):

    knn = KNeighborsClassifier(n\_neighbors=i)

    knn.fit(X\_train,y\_train)

    pred\_i = knn.predict(X\_test)

    error\_rate.append(np.mean(pred\_i != y\_test))

**ERROR RATE VS K VALUE:**

plt.figure(figsize=(10,6))

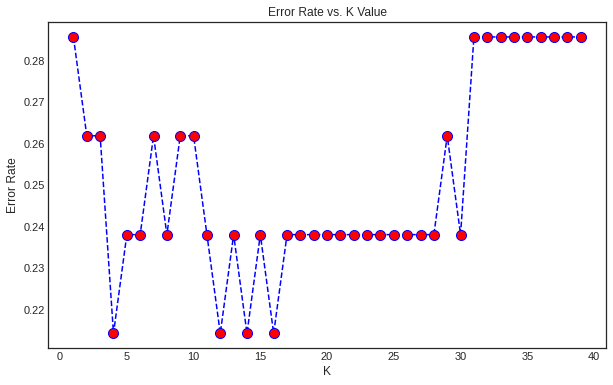
plt.plot(range(1,40),error\_rate,color='blue', linestyle='dashed', marker='o',

         markerfacecolor='red', markersize=10)

plt.title('Error Rate vs. K Value')

plt.xlabel('K')

plt.ylabel('Error Rate')



from sklearn.metrics import confusion\_matrix

knn = KNeighborsClassifier(n\_neighbors=5, metric='euclidean')

knn.fit(X\_train, y\_train)

y\_pred = knn.predict(X\_test)

confusion\_matrix = confusion\_matrix(y\_test, y\_pred)

print("Confusion Matrix:\n",confusion\_matrix)

print("Classification Report:\n",classification\_report(y\_test, y\_pred))

**NAIVE BASE CLASSIFIER:**

#Importing and fitting

from sklearn.naive\_bayes import BernoulliNB

from sklearn.model\_selection import cross\_val\_score

gnb = BernoulliNB()

gnb.fit(X\_train, y\_train)

#Applying and predicting

y\_pred = gnb.predict(X\_test)

cv\_scores = cross\_val\_score(gnb, X, y,

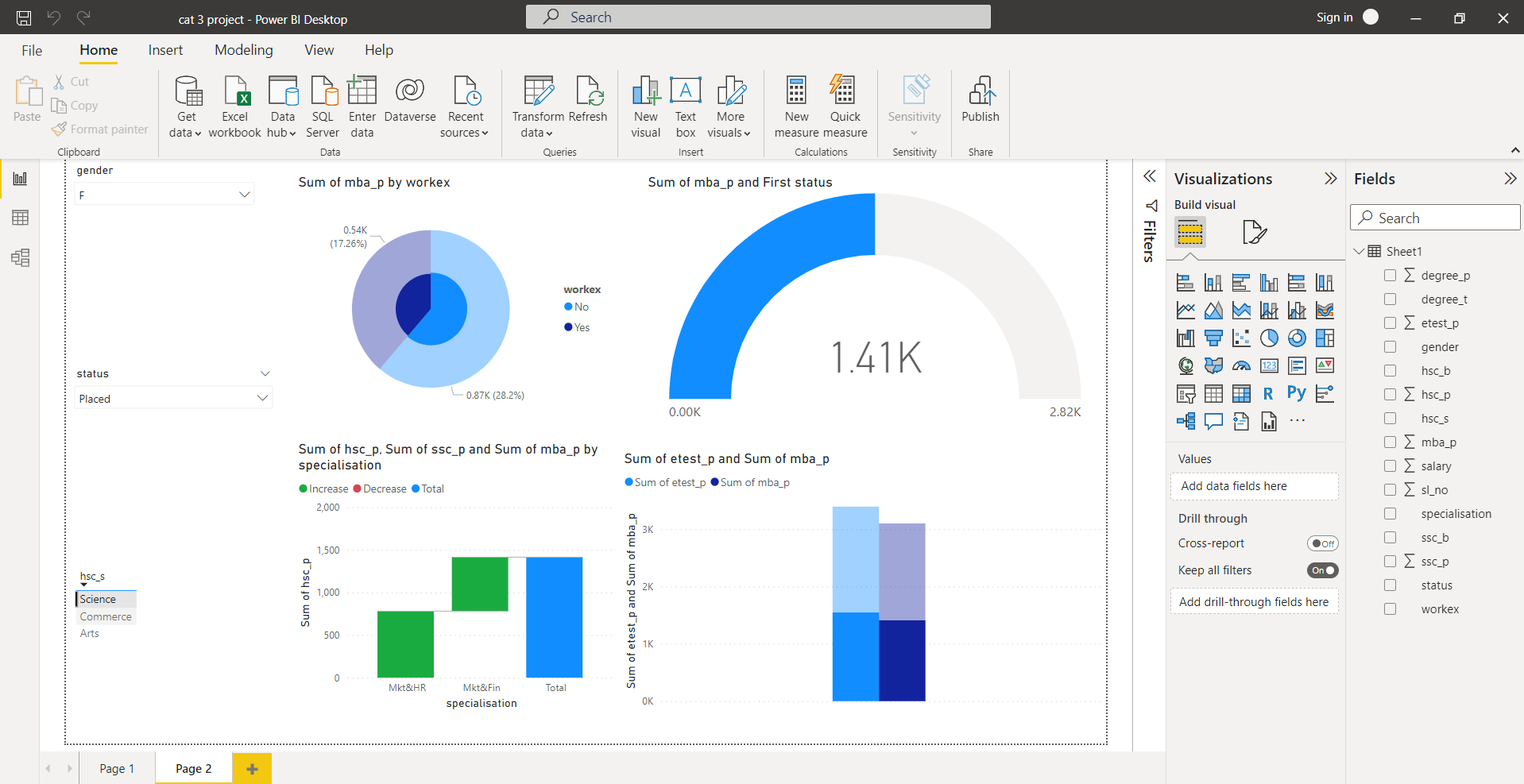
                            cv=10,

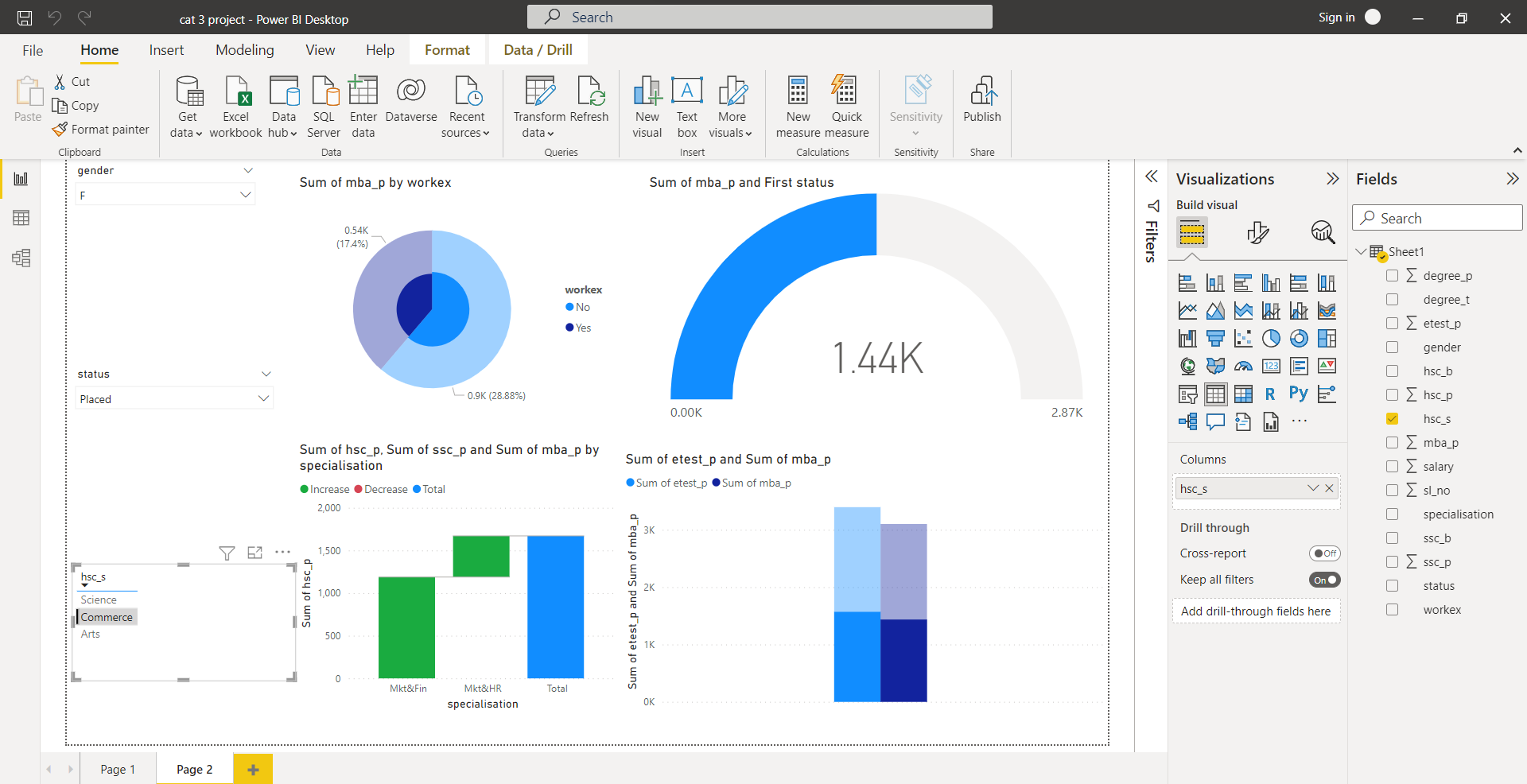
                            scoring='precision')

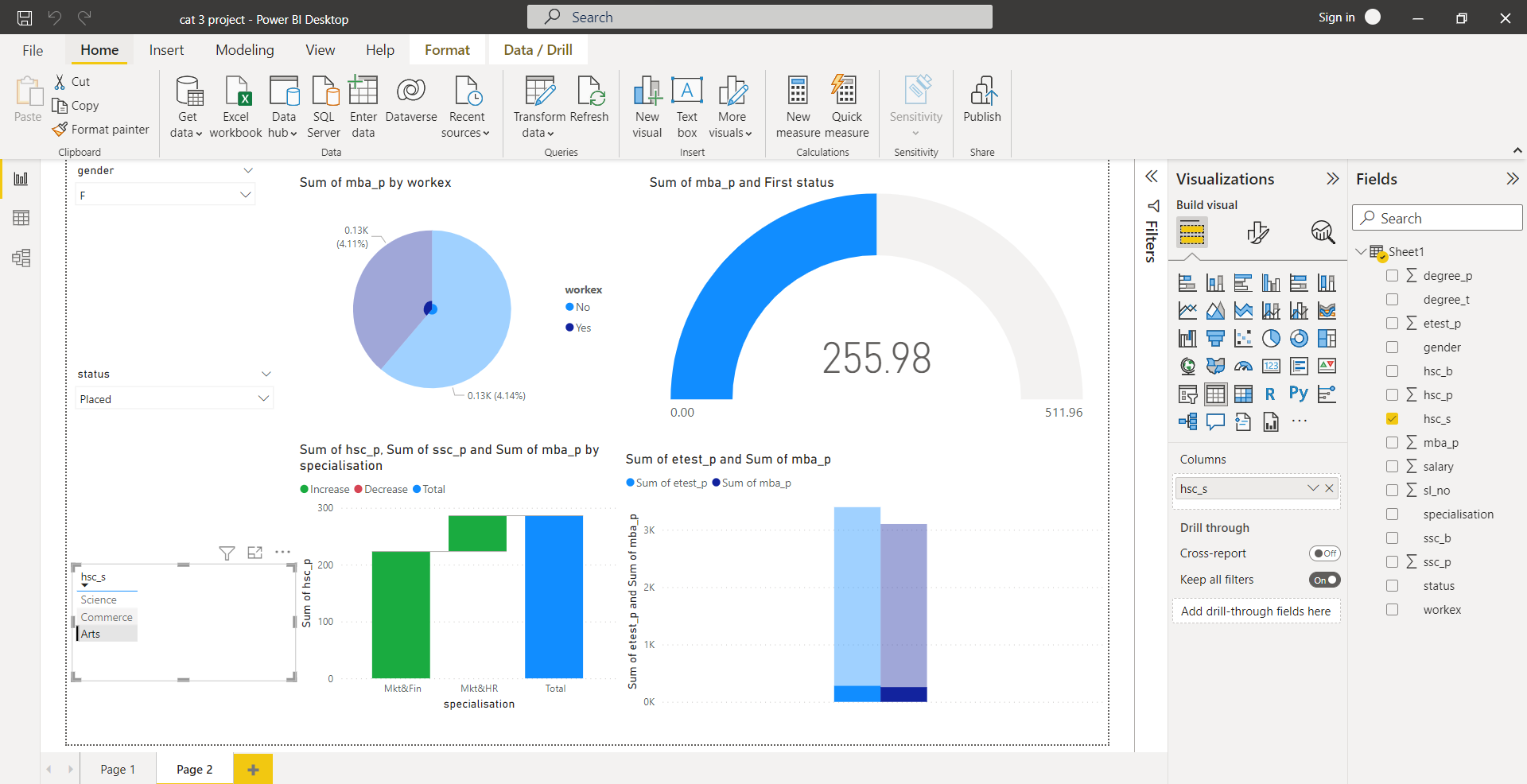
print("Cross-validation precision: %f" % cv\_scores.mean())

ACCURACY: 73.5%

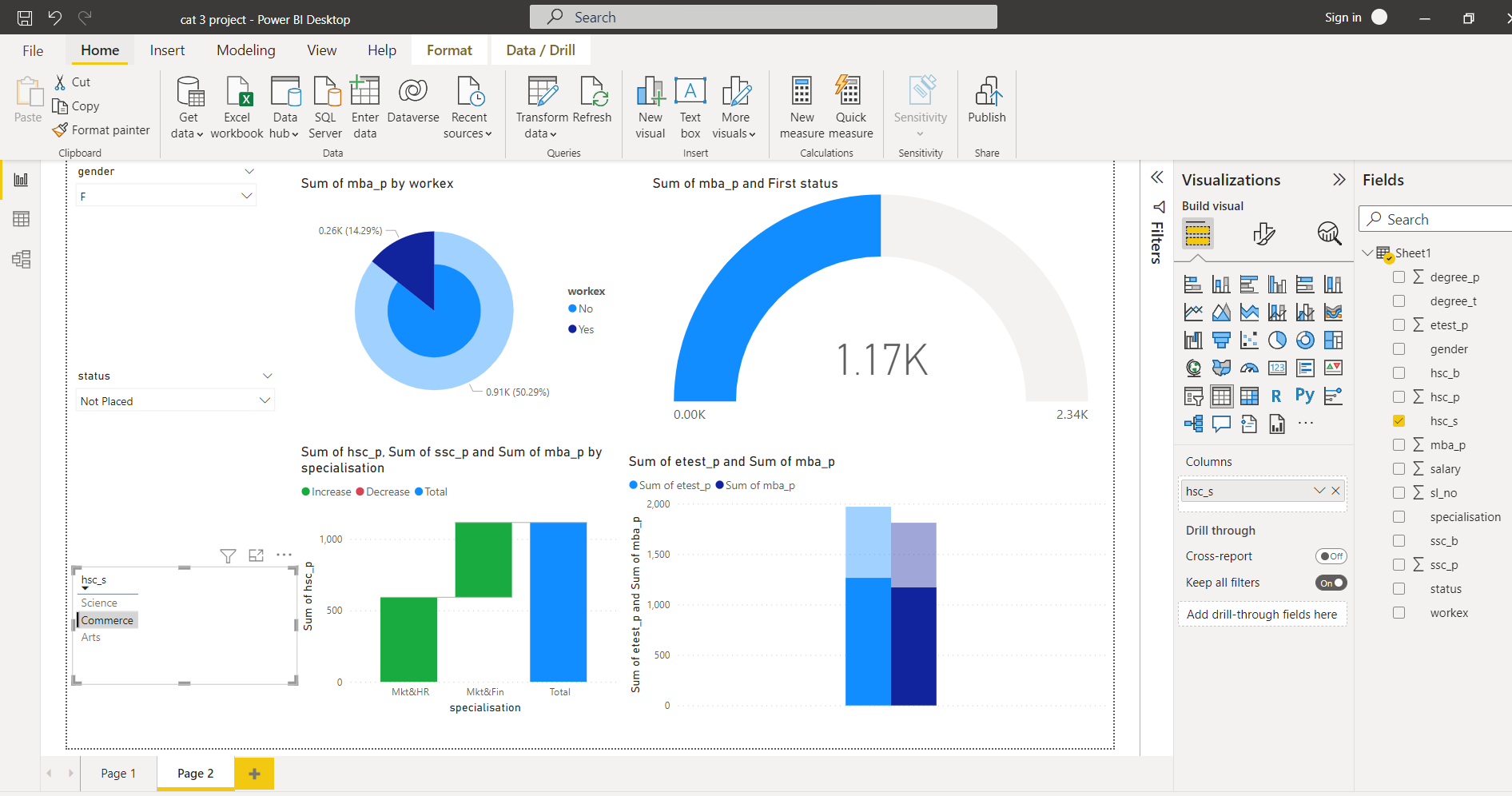
POWER BI:

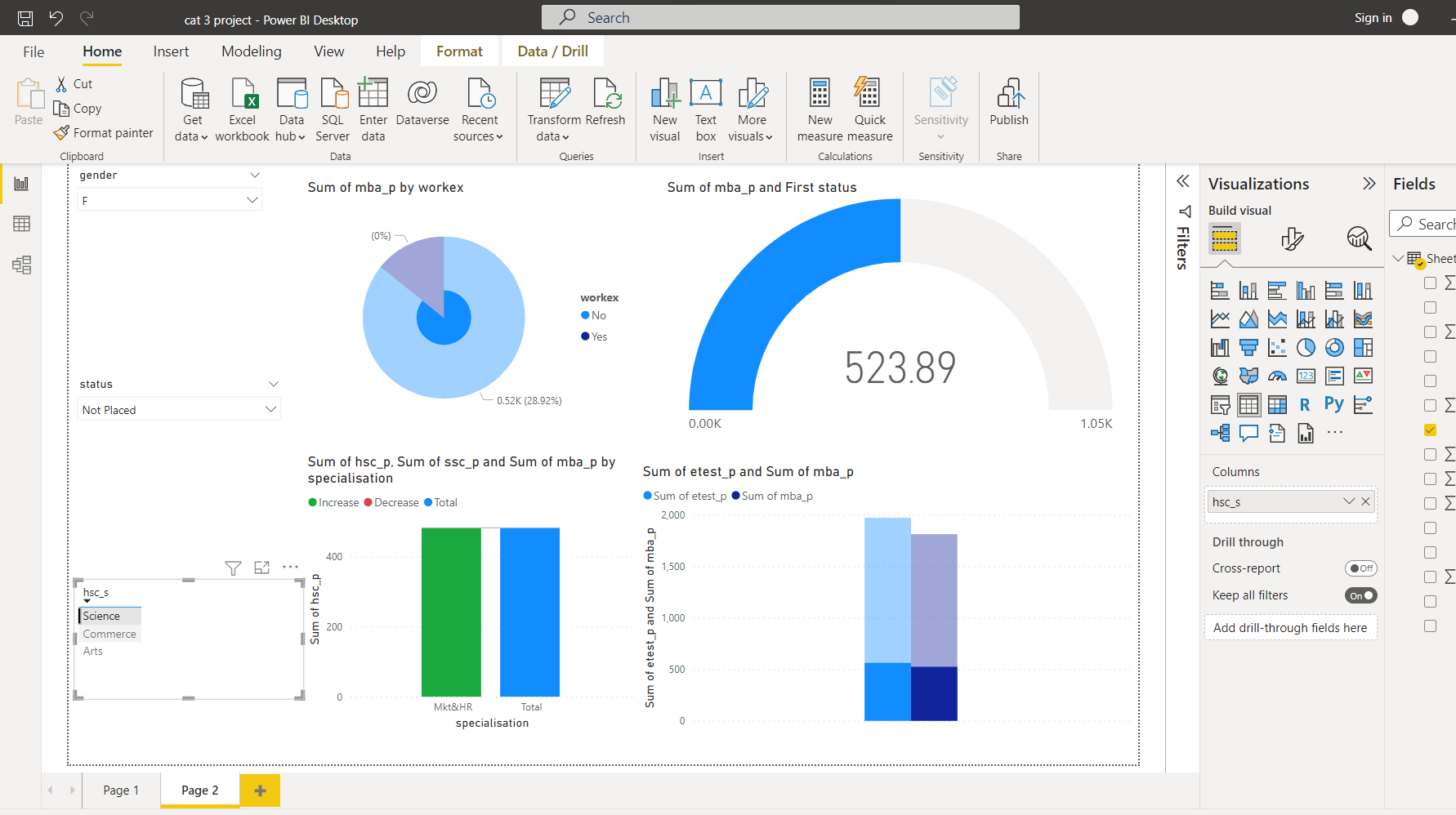


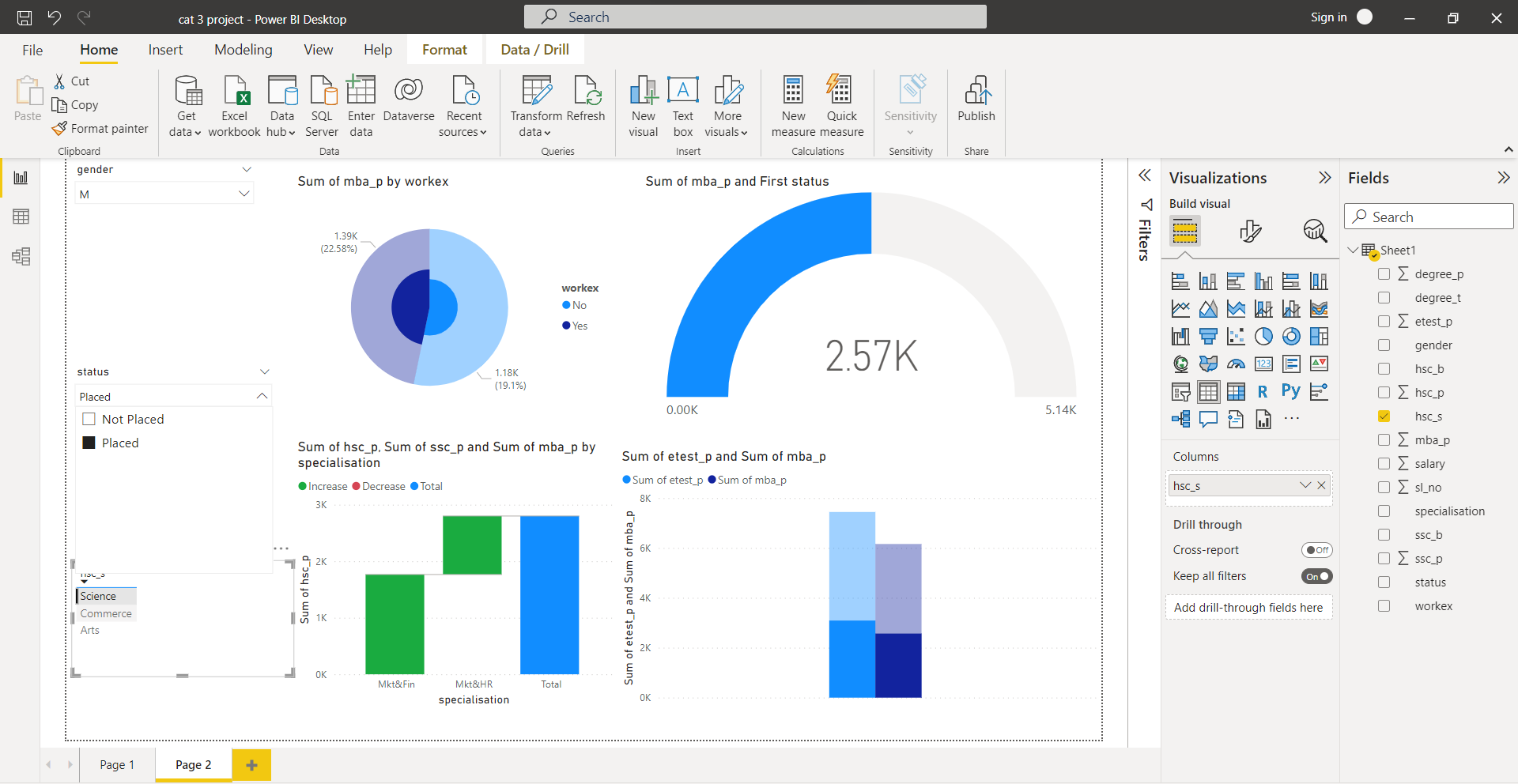


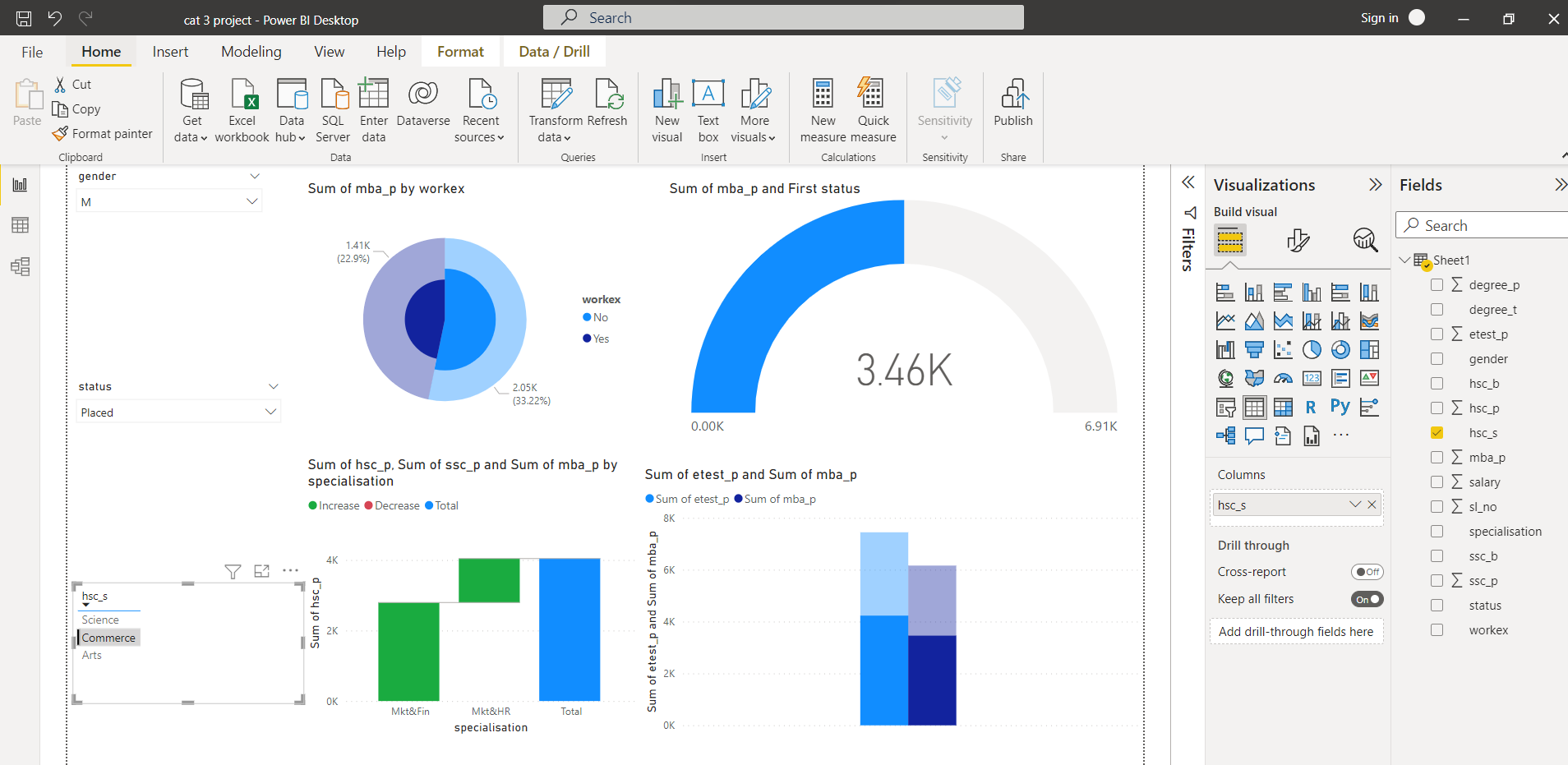


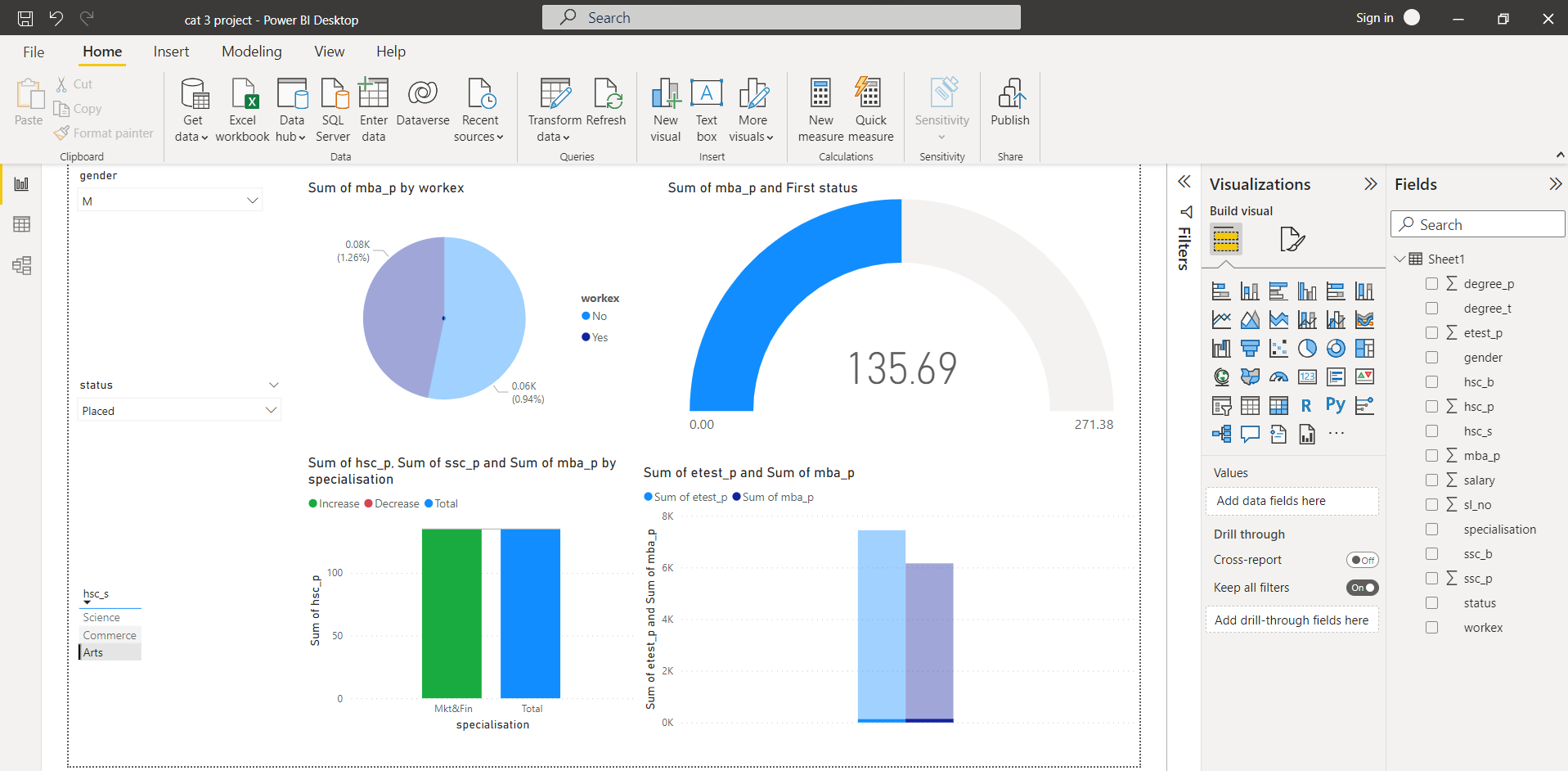




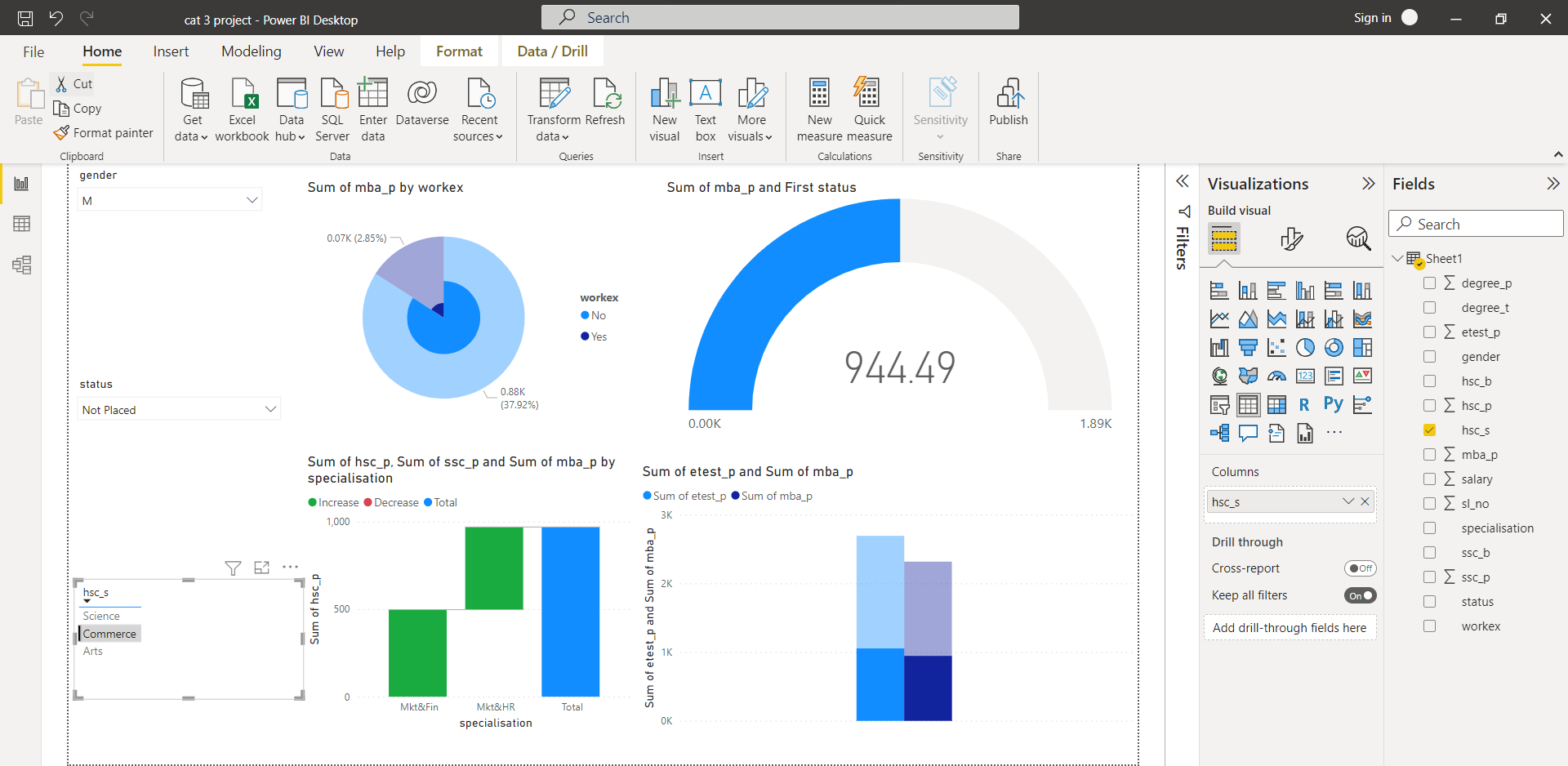


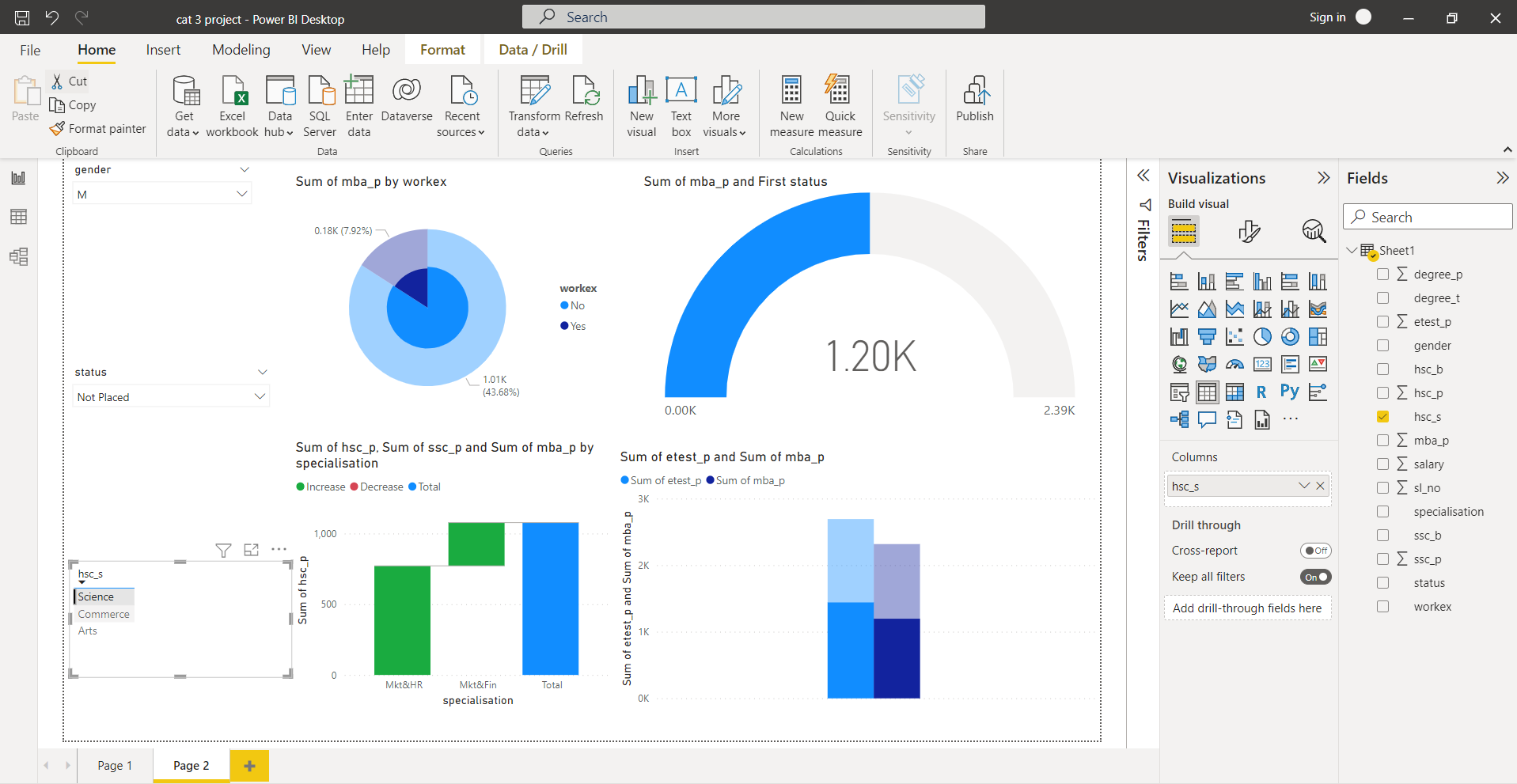


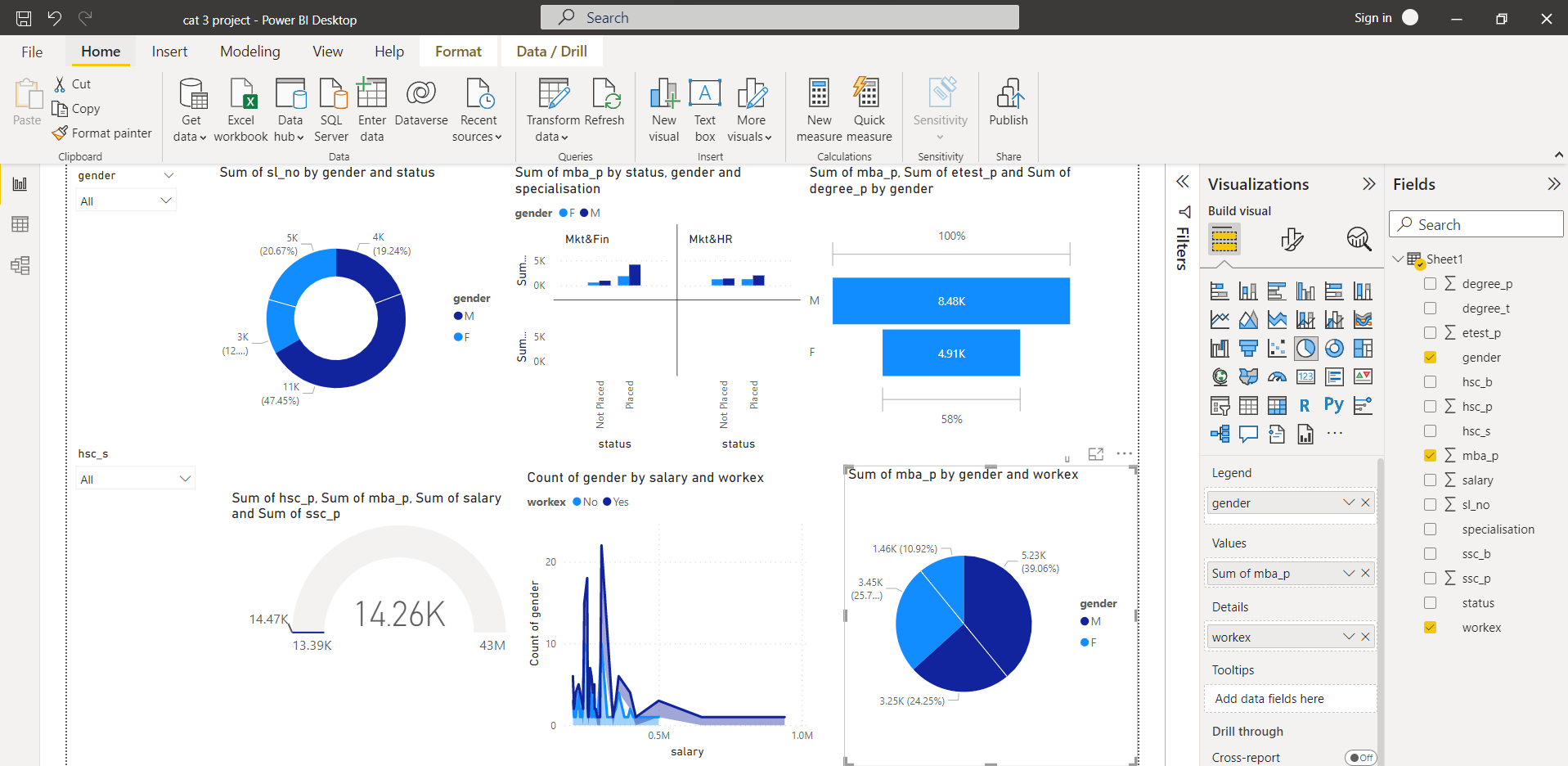


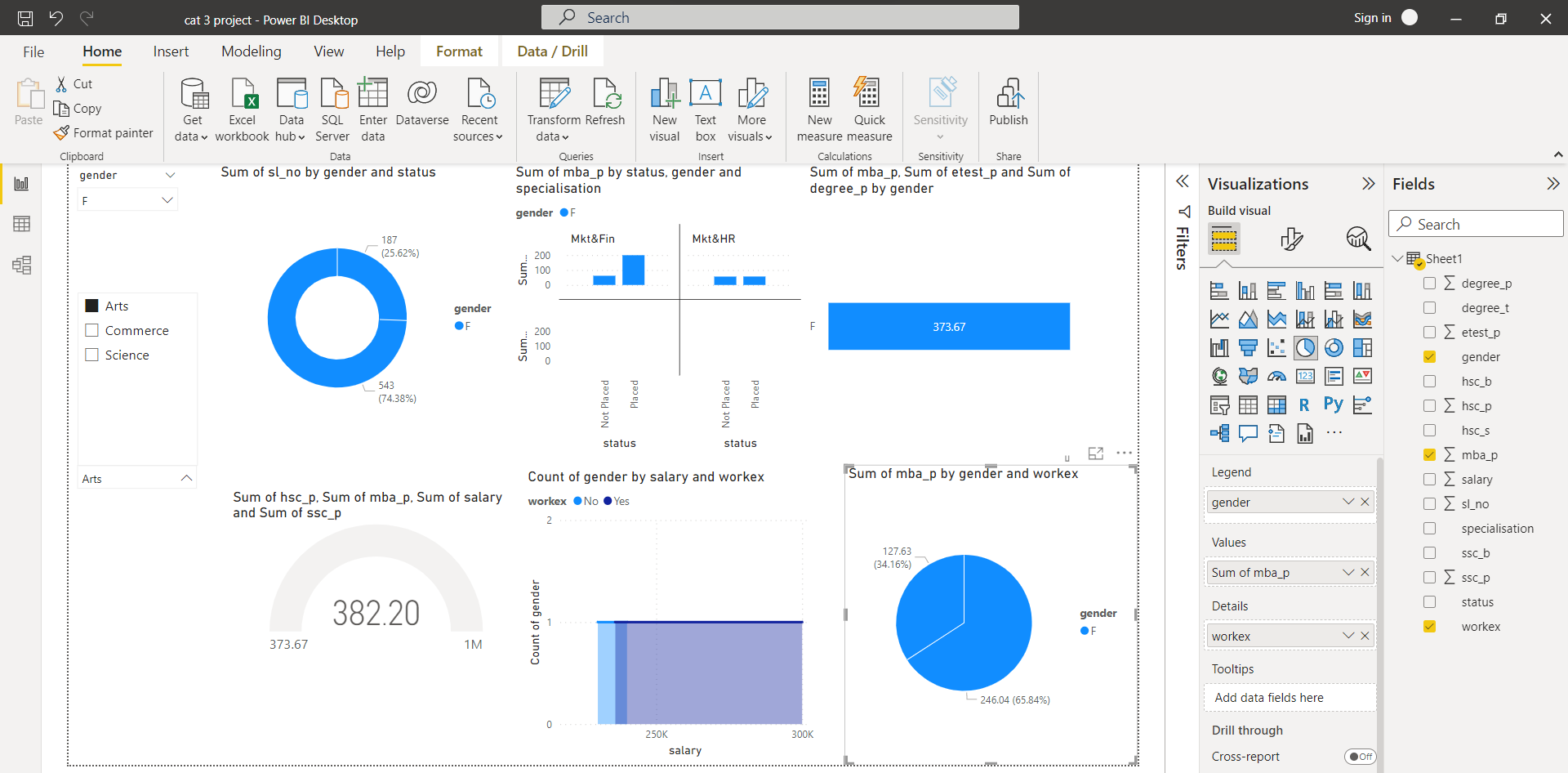


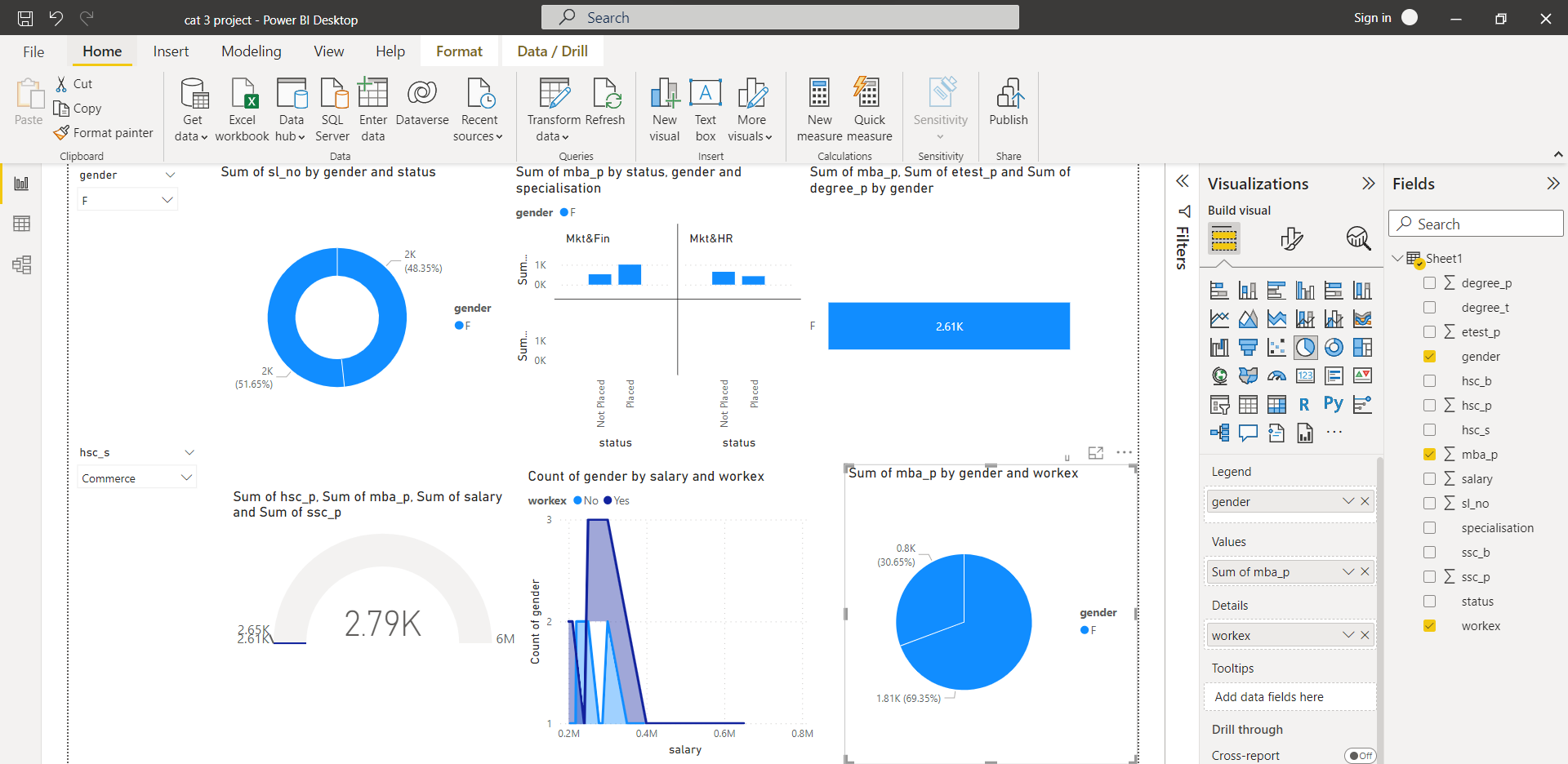


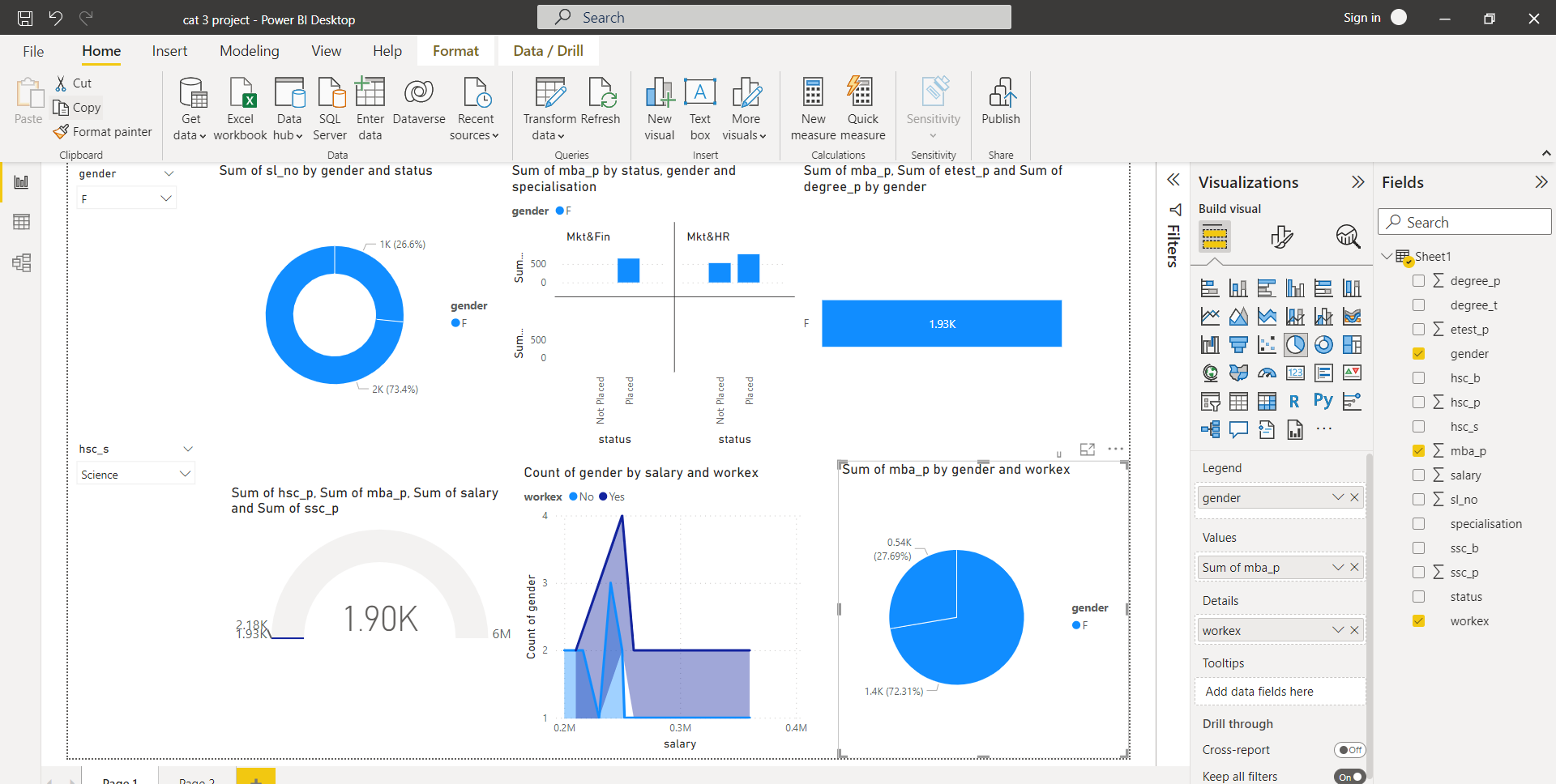


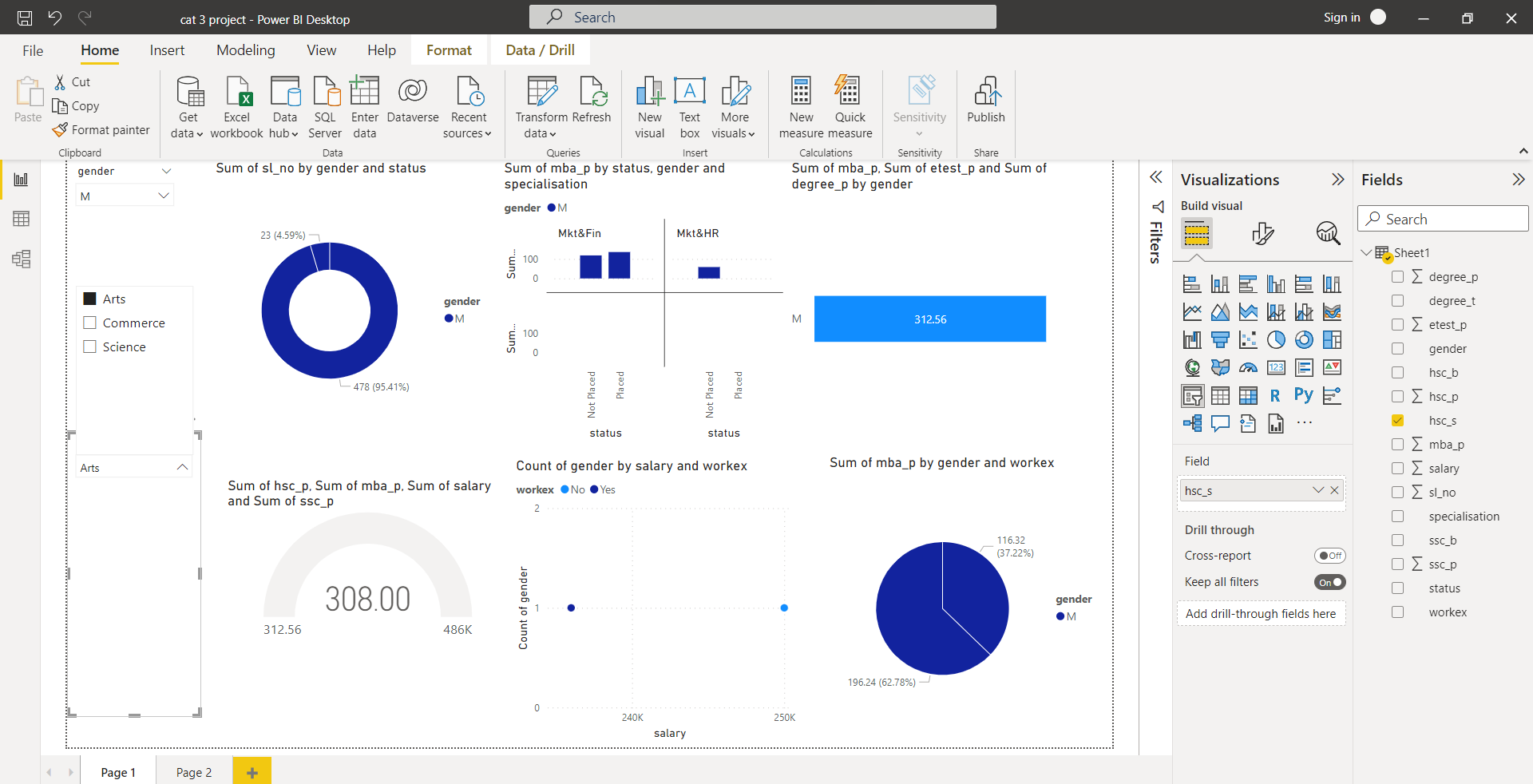


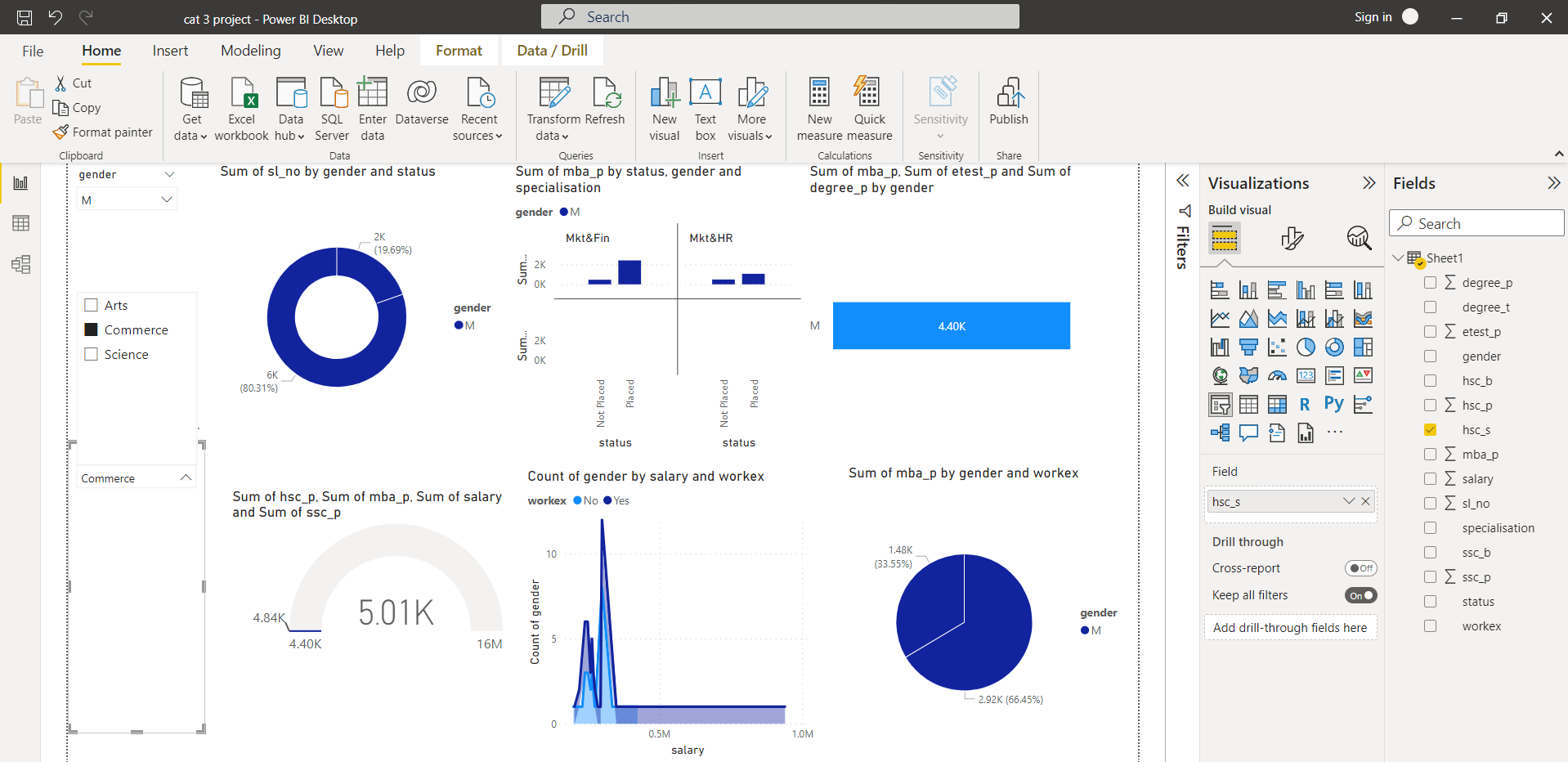


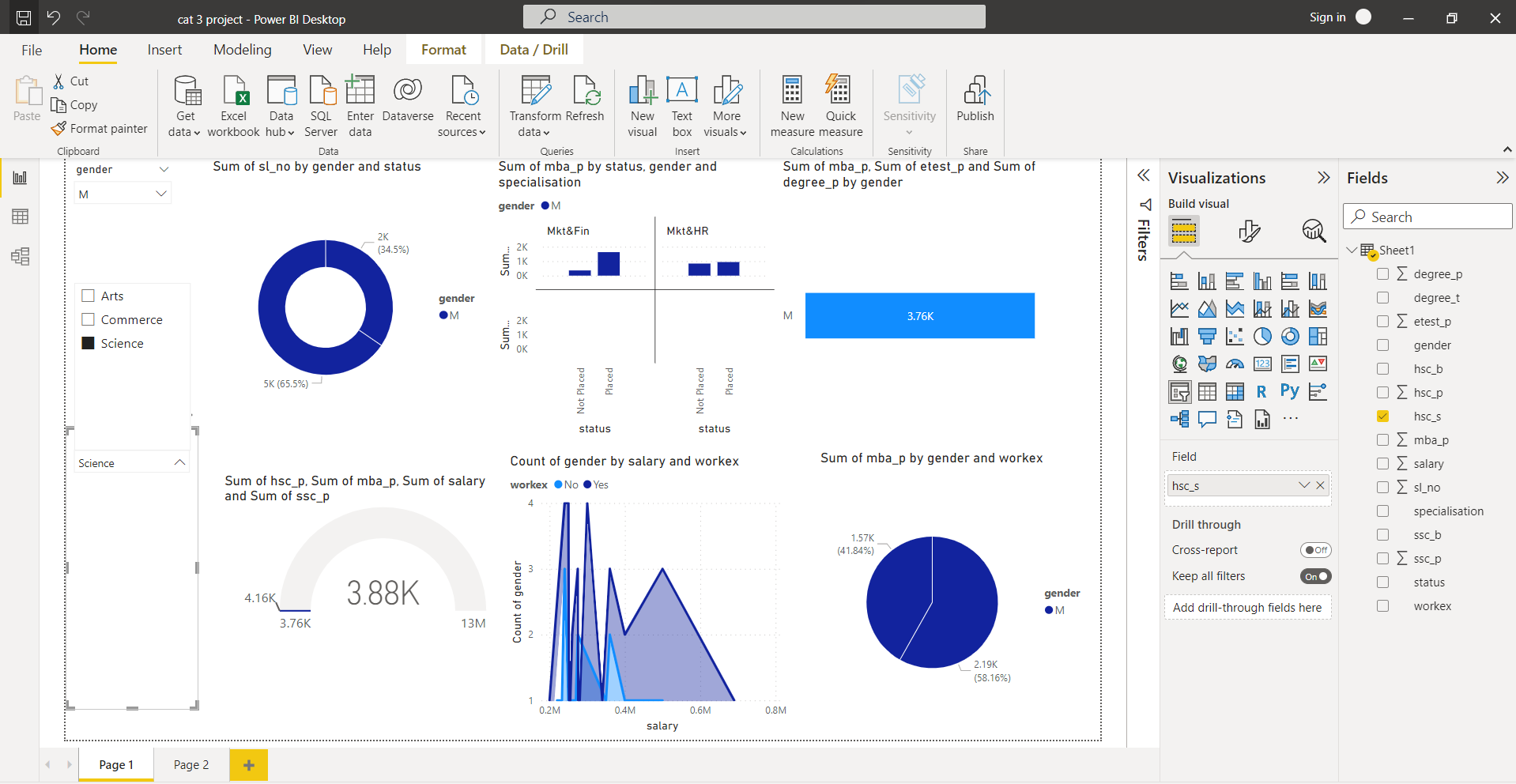












**FINAL SUMMARY:**

**From the analysis report on Campus Recruitment dataset:**

* Educational percentages are highly influential for a candidate to get placed
* Past work experience doesn't influence much on your master final placements
* There are no gender discrimination while hiring, but higher packages were given to male
* Academic percentages have no relation towards salary package.

**Male**

1. Male got not placed over 34.5% and 65.5% got placed.
2. Male got placed in more market& finance and market& human resource.
3. Male has a 3.76k value
4. Male has got placed 58.16% placed without work experience and remaining got placed in work experience.

**Female**

1. Female got placed 73.4% and 27% not placed on science.
2. Female has placed less salary than the work experience that they experienced.