https://gist.github.com/davidemig/eb0b51abb25d9fb87fa736833631f10e

https://github.com/dennislamcv1/IBMDAProject

https://github.com/andyocnz/IBM-Data-Analyst

https://gist.github.com/darryllamoureux/6f57a5e0720df2bc3a99135e0fb50e9e

**Graded Quiz: Collecting Data**

1.Вопрос 1

Which python module helps you to easily access an API?

NumPy

**Requests**

Pandas

Matplotlib

2.Вопрос 2

In the output of GitHub Jobs API, which field name will help you to identify the location of the job?

lat\_long

city

**location**

area

3.Вопрос 3

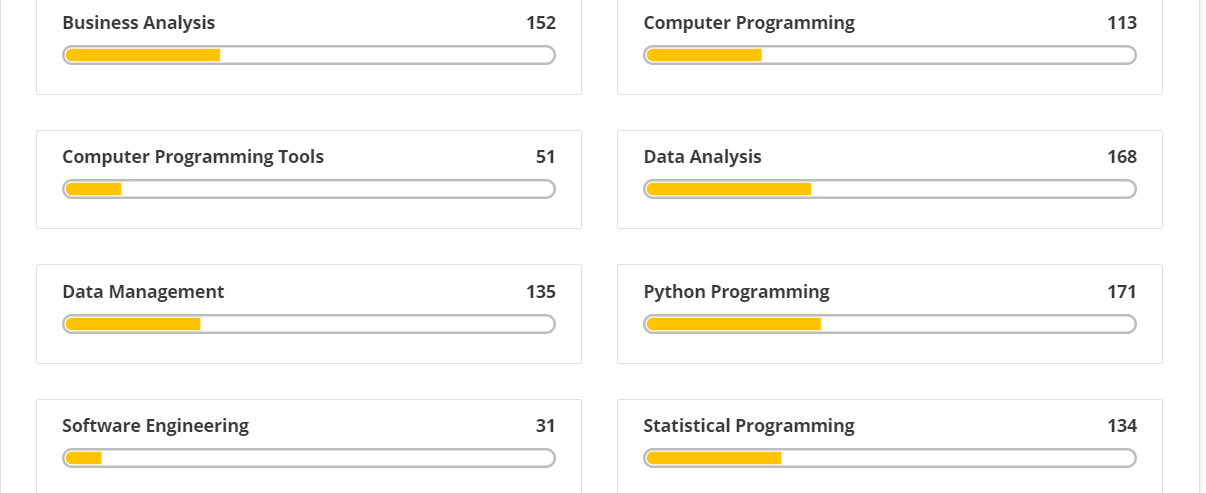
In the output of GitHub Jobs API, which field name will help you to identify if a job is full time or part time?

type\_of\_job

job\_category

jobtype

**type**



**Graded Quiz: Web Scraping**

1.Вопрос 1

Which language are developers paid the most according to the output of the web scraping lab?

**Swift**

Java

C++

C

2.Вопрос 2

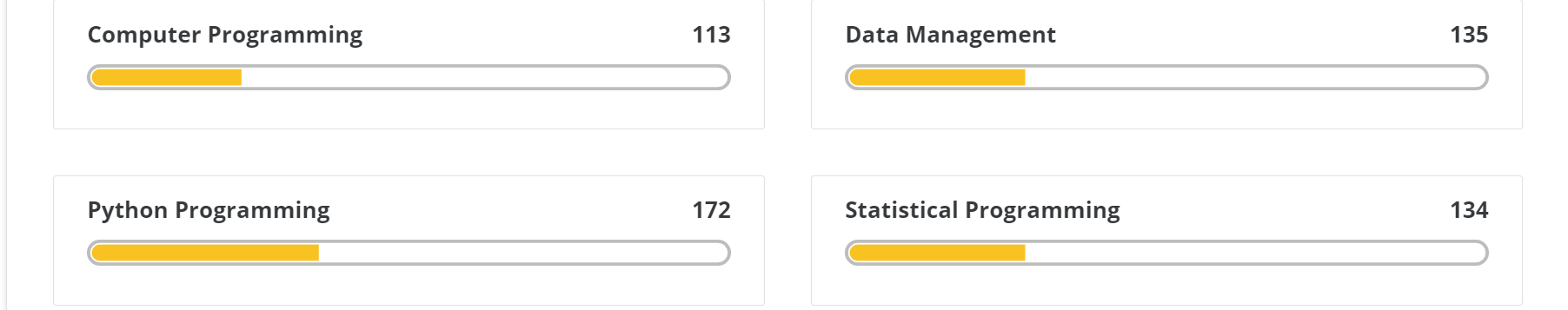
Which of the following code segments will help to scrape all images from a html page?

soup.findall(“pic”)

soup.find\_all(“image”)

soup.findall(“img”)

**soup.find\_all(“img”)**



# Graded Quiz: Exploring Data

1.Вопрос 1

How many rows are in the data set?

11255

11525

11555

**11552**

2.Вопрос 2

How many columns are in the data set?

86

84

26

**85**

3.Вопрос 3

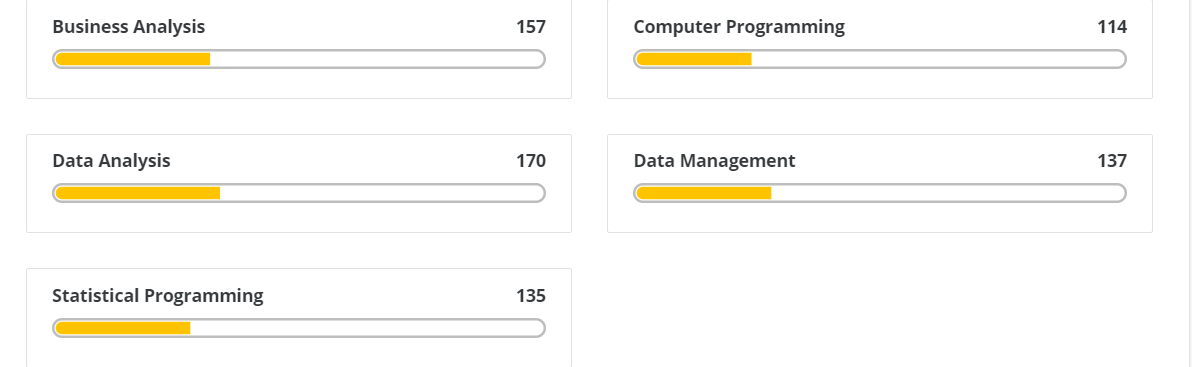
What type of column is "Age"?

**Numerical**

Nominal

Text

Categorical



## Graded Quiz: Duplicates

1.Вопрос 1

How many duplicate rows are there in the dataset?

99

124

**154**

145

2.Вопрос 2

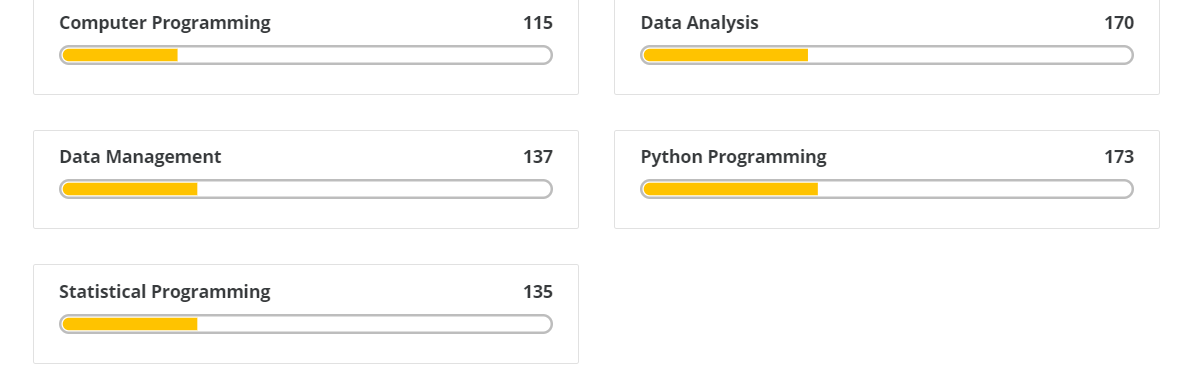
How many duplicate values are there in the column Respondent?

99

**154**

10

96



## Graded Quiz: Removing Duplicates

1.Вопрос 1

After removing the duplicate rows, how many rows are there in the dataset?

**11398**

12333

13456

9999

Правильно

2.Вопрос 2

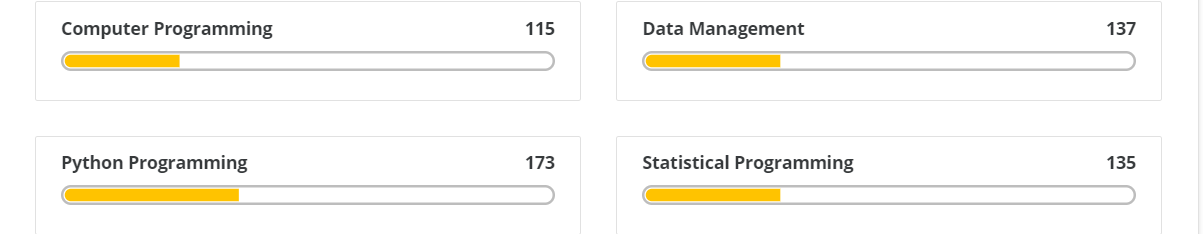
‎После удаления дублирующих строк, сколько уникальных строк есть в столбце Respondent? ‎

**11398**

11342

11999

11000



**Graded Quiz: Missing Values**

1.Вопрос 1

After removing the duplicate rows, how many blank rows are there under the column EdLevel?

280

0

231

**112**

Правильно

2.Вопрос 2

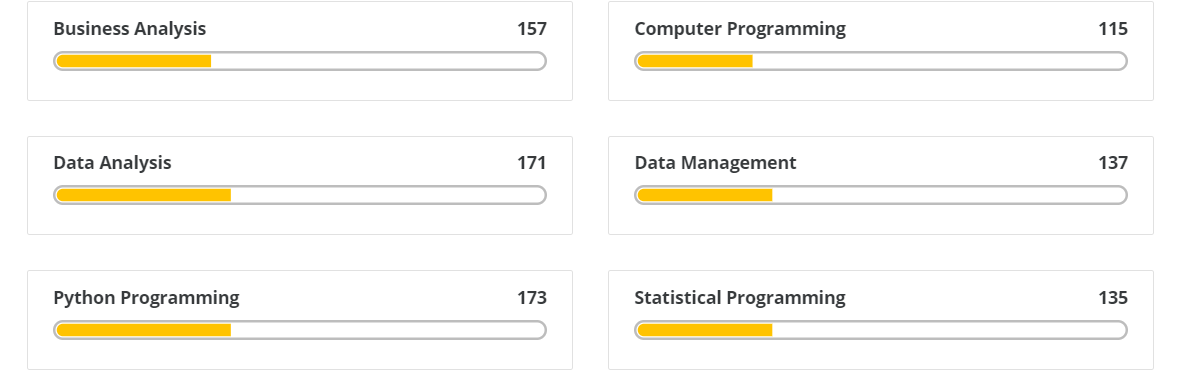
After removing the duplicate rows, how many rows are missing under the column Country?

280

188

201

**0**



**Graded Quiz: Imputing Missing Values**

1.Вопрос 1

What is the majority category under the column Employment?

Independent contractor, freelancer, or self-employed

Retired

Employed part-time

**Employed full-time**

2.Вопрос 2

Under the column " UndergradMajor", which category has the minimum number of rows?

**Health Science**

Fine Arts

Computer Science

Information Systems

3.Вопрос 3

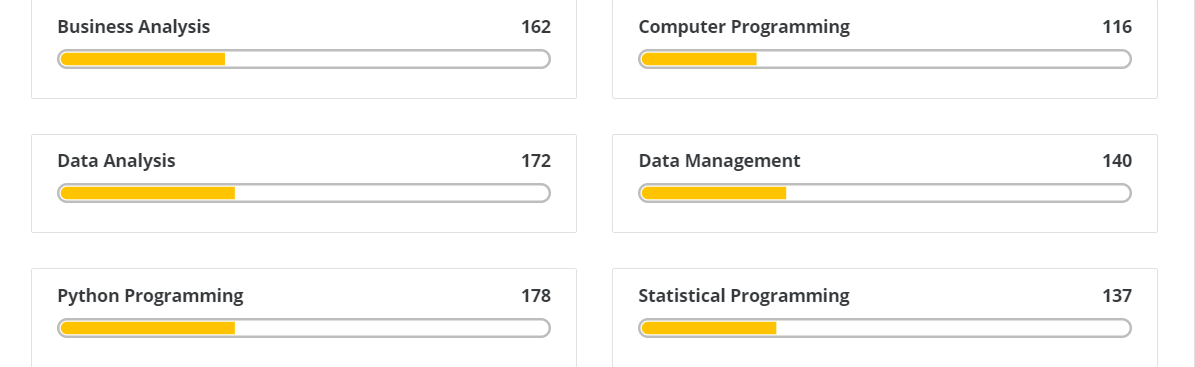
The column ‘ConvertedComp’ contains the annual compensation of the survey respondents. What is the best approach to impute the missing values in this column?

**median**

min

max

mean



**Graded Quiz: Normalizing Data**

1.Вопрос 1

How many unique values are there in the CompFreq column?

**3**

11398

5

7

2.Вопрос 2

After removing the duplicate rows, how many respondents are being paid yearly?

312

**6093**

3259

9999

3.Вопрос 3

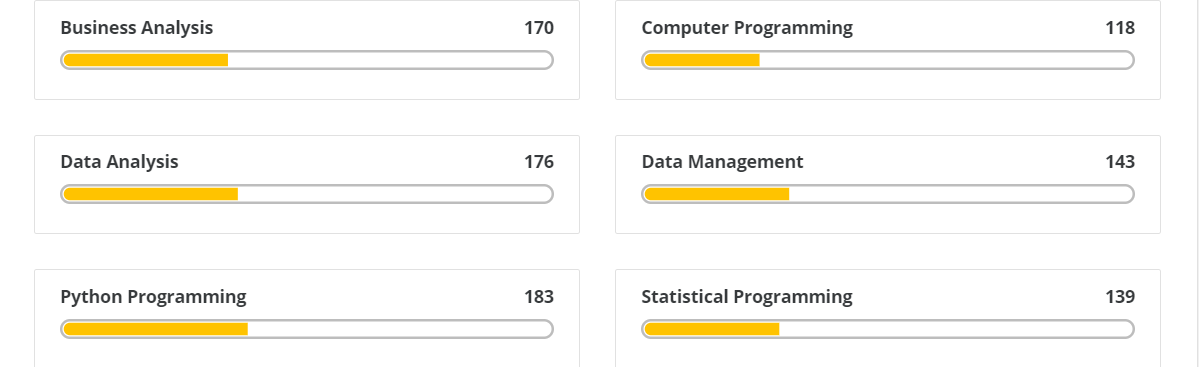
What is the median NormalizedAnnualCompensation?

9000

6132520

8080

**100000**



# Graded Quiz: Distribution

1.Вопрос 1

What is the median Age of survey respondents?

31

**29**

28

30

2.Вопрос 2

What is the median ConvertedComp of respondents who have identified themselves as ‘Woman’?

**57708**

54678

55467

56453

3.Вопрос 3

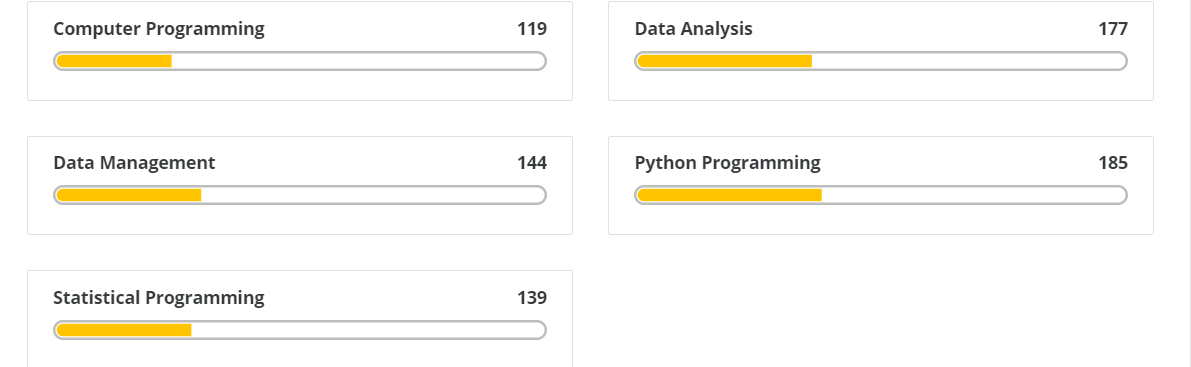
Based on the histogram, which age group has the highest number of survey respondents?

**24-32**

16-24

32-40

40-48



# Graded Quiz: Handling Outliers

1.Вопрос 1

What is the median ConvertedComp before removing outliers?

78675

**57745**

55342

55768

2.Вопрос 2

What is the median ConvertedComp after removing outliers?

56456

53564

**52704**

55672

3.Вопрос 3

Based on the boxplot of ‘Age’ how many outliers do you see below Q1?

3

2

1

**0**

4.Вопрос 4

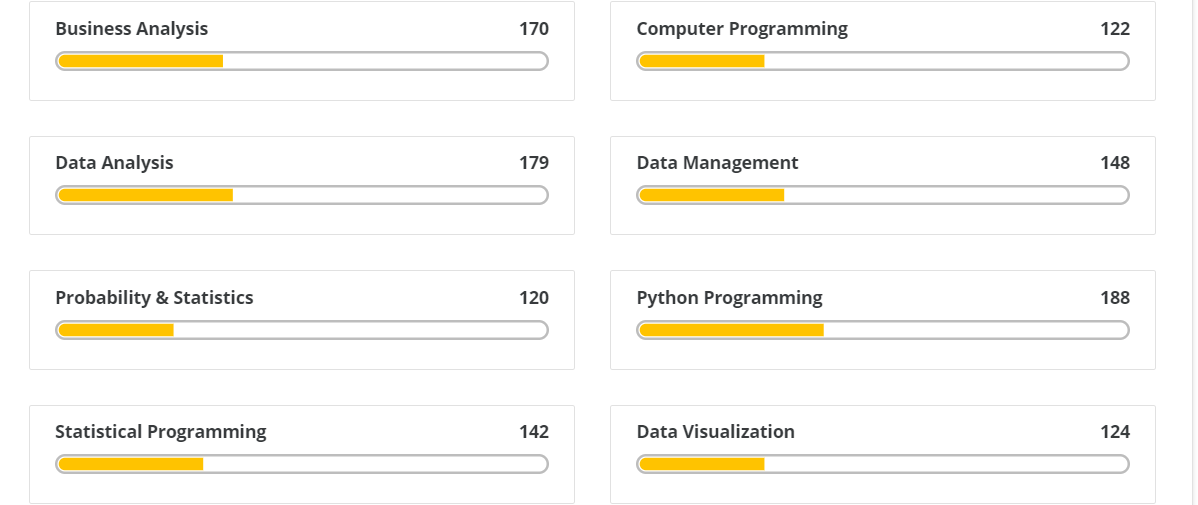
What is the mean ConvertedComp after removing outliers?

2378.56

34567.34

**59883**

10000



**Graded Quiz: Correlation**

1.Вопрос 1

Which column has a negative correlation with "Age"?

WorkWeekHrs

CompTotal

**CodeRevHrs**

Respondent

2.Вопрос 2

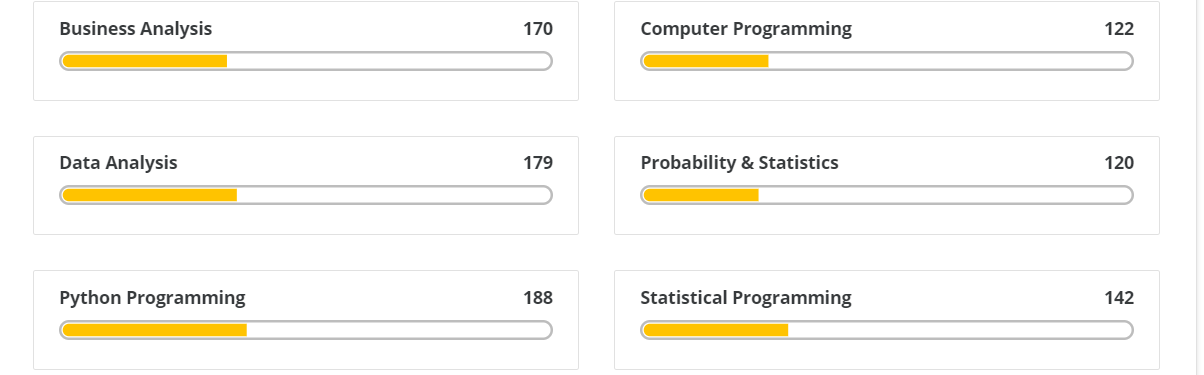
Which column has the highest correlation with "Age"?

Respondent

CompTotal

**ConvertedComp**

WorkWeekHrs



**W4**

**Graded Quiz: Visualizing Distribution Of Data**

1.Вопрос 1

Based on the boxplot of the column "Age", where do you see the outliers?

Below Q1

**Above Q3**

There are no outliers

Above Q3 and Below Q1

**Graded Quiz: Visualizing Relationship**

1.Вопрос 1

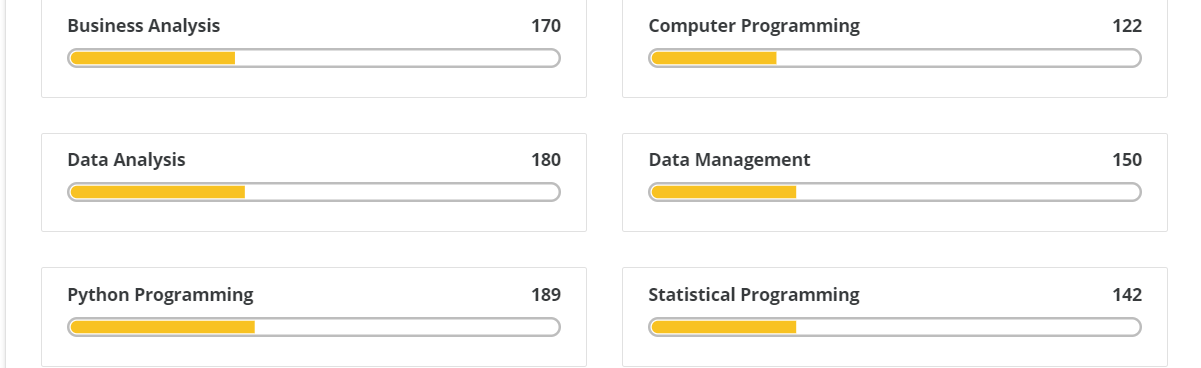
Based on the scatter plot of Age and WorkWeekHrs , what kind of correlation do you see between them?

Very high positive correlation

**Difficult to clearly identify from the scatter plot.**

No correlation

Very High negative correlation.



**Graded Quiz: Visualizing Composition of Data**

1.Вопрос 1

In the pie chart of top 5 databases, respondents wish to learn next year, what is the percentage of Mongodb?

18.8

19.1

16.4

**20.9**

2.Вопрос 2

In the list of most popular languages respondents wish to learn next year, what is the rank of Python?

**3**

2

4

1

3.Вопрос 3

How many respondents indicated that they currently work with 'SQL'?

**7106**

5132

5012

5321

4.Вопрос 4

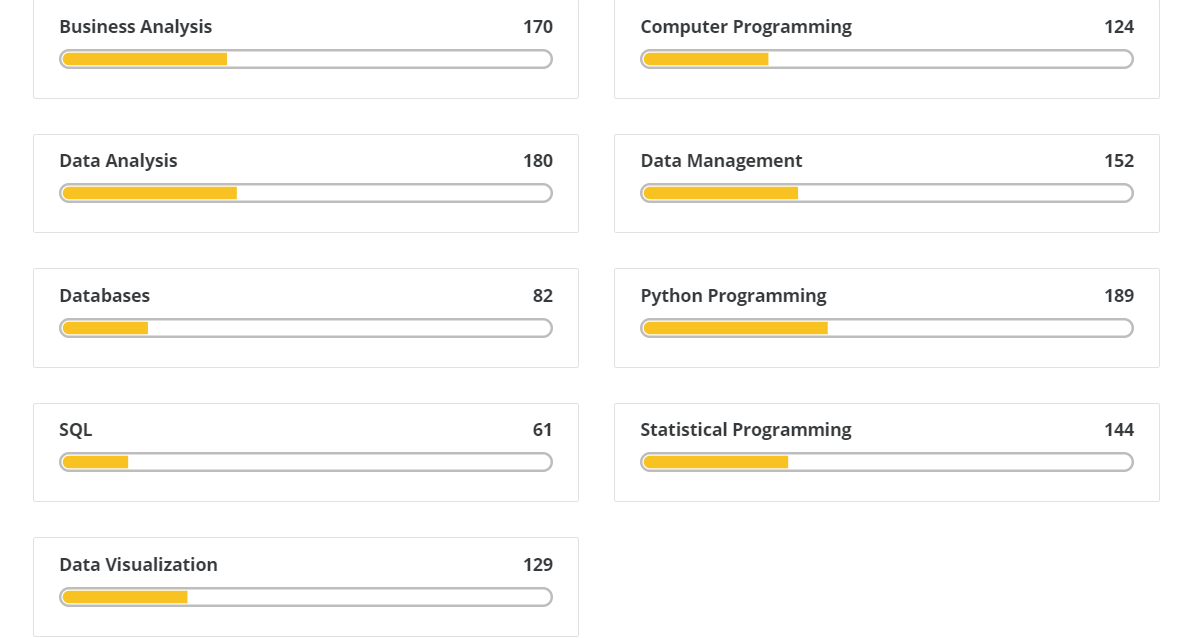
How many respondents indicated that they work on 'MySQL' only?

**474**

736

657

876



**Graded Quiz: Visualizing Comparison of Data**

1.Вопрос 1

‎Большинство респондентов, ответив на опрос,‎

Developer - front-end

**Developer - back-end**

Developer - full-stack

Developer - mobile

2.Вопрос 2

Based on the Line Chart of Age and median ConvertedComp for the age group 25-30, what kind of correlation do you see between Age and median ConvertedComp?

**Positive**

No Correlation

Negative

Highly Negative

