

사람과의 대화처럼 다음 그림은 파이썬과 대화를 나누는 모습입니다. 크게 세 개의 질문과 대답 쌍으로 이루어져 있습니다. 첫 번째 질문은 1+2는 무엇인지, 두 번째 질문은 2\*3\*4\*5는 무엇인지, 세 번째 질문은 '안녕'이라고 작성한 경우에 각각의 답변을 확인할 수 있습니다. 별다른 기호나 명령어를 입력하지 않더라도 직관적인 질문과 대답을 이어갈 수 있습니다.

이러한 대화식의 환경은 파이썬을 설치하는 경우에 기능을 지원하고 있고, 일반적으로 사용하는 주피터 노트북(Jupyter Notebook)이나 파이참(PyCharm), 아나콘다(Anaconda) 등에서 모두 사용할 수 있습니다. 여기에서는 설명의 편의를 위해 파이썬을 아나콘다 콘솔 환경에서 진행하며, >>> 뒤에 작성하는 내용은 우리가 직접 작성하는 질문이고, 바로 아래 출력되는 내용이 파이썬이 대답하는 부분임을 참고하기 바랍니다.

```

>>> 1+2
3
>>> 2*3*4*5
120
>>> '안녕'
'안녕'

```

파이썬과의 대화

### 1.2 파이썬과 약속하기

매우 쉽고 자유로운 파이썬이지만, 몇 가지 규칙들이 있습니다. 이러한 기본 규칙들을 이해해야 앞으로 남은 학습 과정을 순조롭게 진행할 수 있습니다.

## 229p 코드 오류 수정 (vs 칼럼에서 중복 제거한 값을 구하기)

시다. 중복 제거한 값들을 확인하기 위해서는 `unique()` 함수를 사용합니다. 다음 결과와 같이 `am` 열은 `manual`, `auto`라는 2개의 값으로 구성되어 있고, `gear` 열은 3, 4, 5, \*3, \*5라는 5개의 값, `vs` 열은 2개의 값으로 구성되어 있음을 알 수 있습니다.

여기서 전진기어 개수를 의미하는 `gear` 열에는 숫자가 아닌 \*3, \*5가 포함되어 있고, 엔진 유형을 의미하는 `vs` 열에는 0, 1이라는 값만 포함되어 있음을 확인할 수 있습니다. 이러한 문제는 나중에 전처리 단계를 통해서 가공할 예정입니다.

```
>>> print(data['am'].unique())           # am 칼럼에서 중복 제거한 값들 구하기
['manual' 'auto']
>>> print(data['gear'].unique())         # gear 칼럼에서 중복 제거한 값들 구하기
['4' '3' '*3' '5' '*5']
>>> print(data['vs'].unique())           # vs 칼럼에서 중복 제거한 값들 구하기
[0 1]
```

### ■ 요약정보 확인하기: info

데이터의 **요약정보**를 확인하기 위해서 `info()` 함수를 사용합니다. 요약정보를 통해 대략적으로 범주형 변수와 연속형 변수를 예상할 수 있고, 전체 데이터 크기, Null 값의 개수 등을 파악할 수 있습니다.

다음 결과를 통해서 전체 데이터 건수는 32건(32 entries)이고, 구성된 칼럼의 개수는 12개(total 12 columns)라는 사실을 재차 확인할 수 있습니다. 또한 `cyl` 열의 Non-Null Count 값이 30 non-null인 것으로 보아 2개의 결측치가 존재하며, `qsec` 열에도 1개의 결측치가 있다는 사실을 알 수 있습니다.

이번에는 데이터 타입(Dtype은 Data type의 약자) 정보를 확인해봅시다. 그전에 범주형 변수의 데이터 타입은 `object` 또는 `string` 등으로 출력되고, 연속형 변수의 데이터 타입은 `int64` 또는 `float64`로 출력된다는 사실을 알아둡시다. 단, 범주형 변수의 값들이 문자열이라면 보통 자동으로 `object` 타입으로 인식되나, 숫자로 된 범주형 변수는 `int64` 등으로 인식될 수 있습니다.