

[Home](#) [Modules](#) [Agents](#) [Toolkits](#) [CSV Agent](#)

# CSV Agent

This notebook shows how to use agents to interact with a csv. It is mostly optimized for question answering.

**NOTE: this agent calls the Pandas DataFrame agent under the hood, which in turn calls the Python agent, which executes LLM generated Python code - this can be bad if the LLM generated Python code is harmful. Use cautiously.**

```
from langchain.agents import create_csv_agent
```

```
from langchain.llms import OpenAI
from langchain.chat_models import ChatOpenAI
from langchain.agents.agent_types import AgentType
```

## Using ZERO\_SHOT\_REACT\_DESCRIPTION #

---

This shows how to initialize the agent using the ZERO\_SHOT\_REACT\_DESCRIPTION agent type. Note that this is an alternative to the above.

```
agent = create_csv_agent(
    OpenAI(temperature=0),
    "titanic.csv",
    verbose=True,
```

```
agent_type=AgentType.ZERO_SHOT_REACT_DESCRIPTION,  
)
```

## Using OpenAI Functions

---

This shows how to initialize the agent using the OPENAI\_FUNCTIONS agent type. Note that this is an alternative to the above.

```
agent = create_csv_agent(  
    ChatOpenAI(temperature=0, model="gpt-3.5-turbo-0613"),  
    "titanic.csv",  
    verbose=True,  
    agent_type=AgentType.OPENAI_FUNCTIONS,  
)
```

```
agent.run("how many rows are there?")
```

```
Error in on_chain_start callback: 'name'
```

```
Invoking: `python_repl_ast` with `df.shape[0]`
```

```
891There are 891 rows in the dataframe.
```

```
> Finished chain.
```

```
'There are 891 rows in the dataframe.'
```

```
agent.run("how many people have more than 3 siblings")
```

```
Error in on_chain_start callback: 'name'
```

```
Invoking: `python_repl_ast` with `df[df['SibSp'] > 3]['PassengerId'].count()`
```

```
30There are 30 people in the dataframe who have more than 3 siblings.
```

```
> Finished chain.
```

```
'There are 30 people in the dataframe who have more than 3 siblings.'
```

```
agent.run("whats the square root of the average age?")
```

```
Error in on_chain_start callback: 'name'
```

```
Invoking: `python_repl_ast` with `import pandas as pd  
import math
```

```
# Create a dataframe  
data = {'Age': [22, 38, 26, 35, 35]}  
df = pd.DataFrame(data)
```

```
# Calculate the average age  
average_age = df['Age'].mean()
```

```
# Calculate the square root of the average age  
square_root = math.sqrt(average_age)
```

```
square_root`
```

```
5.585696017507576The square root of the average age is approximately 5.59.
```

```
> Finished chain.
```

```
'The square root of the average age is approximately 5.59.'
```

## Multi CSV Example

This next part shows how the agent can interact with multiple csv files passed in as a list.

```
agent = create_csv_agent(  
    ChatOpenAI(temperature=0, model="gpt-3.5-turbo-0613"),  
    ["titanic.csv", "titanic_age_fillna.csv"],  
    verbose=True,  
    agent_type=AgentType.OPENAI_FUNCTIONS,  
)  
agent.run("how many rows in the age column are different between the two dfs?")
```

Error in on\_chain\_start callback: 'name'

Invoking: `python\_repl\_ast` with `df1['Age'].nunique() - df2['Age'].nunique()`

-1There is 1 row in the age column that is different between the two dataframes.

> Finished chain.

'There is 1 row in the age column that is different between the two dataframes.'