

Title: ITMD526_Assignment_02

First Name	Last Name	CWID
Arpit	Khandekar	A20409171

Table of Contents

1. Transformation	2
A) Excel to Mysql DB Transformation	2
B) Filter rows with null values Transformation.....	9

1. Transformation

A) Excel to Mysql DB Transformation

1. First, I have created a table target_db.drivers_info in schema target_db by using following query:
CREATE TABLE target_db.drivers_info
(DriverId INT(12), DriverRef VARCHAR(20), Firstname VARCHAR(20), Surname VARCHAR(20),
Nationality VARCHAR(20)
);

The screenshot shows a SQL query being executed in a database client. The query is:

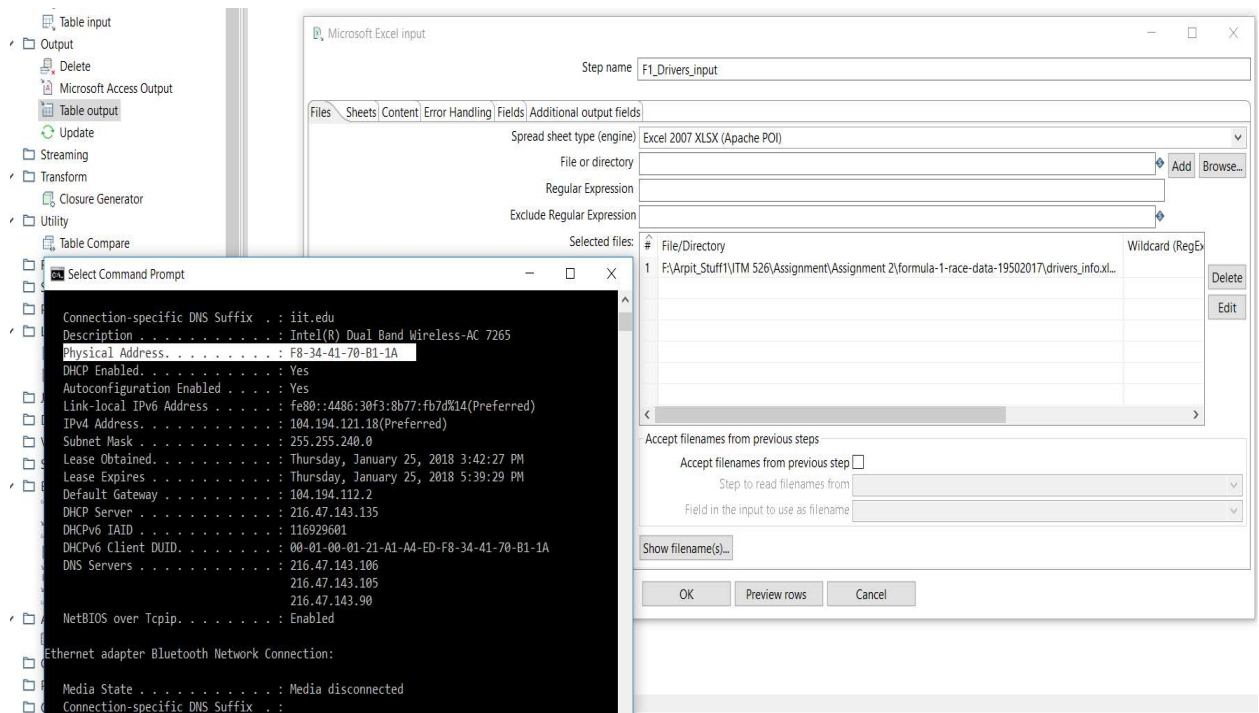
```
USE target_db;
CREATE TABLE target_db.drivers_info
( DriverId INT(12), DriverRef VARCHAR(20), Firstname VARCHAR(20), Surname VARCHAR(20), Nationality VARCHAR(20)
);
SELECT * FROM target_db.drivers_info;
```

The result of the query is shown in a table with the following columns: DriverId, DriverRef, Firstname, Surname, and Nationality.

Next to the SQL query, there is a window titled "Select Command Prompt" showing network configuration details for a Wireless LAN adapter Wi-Fi:

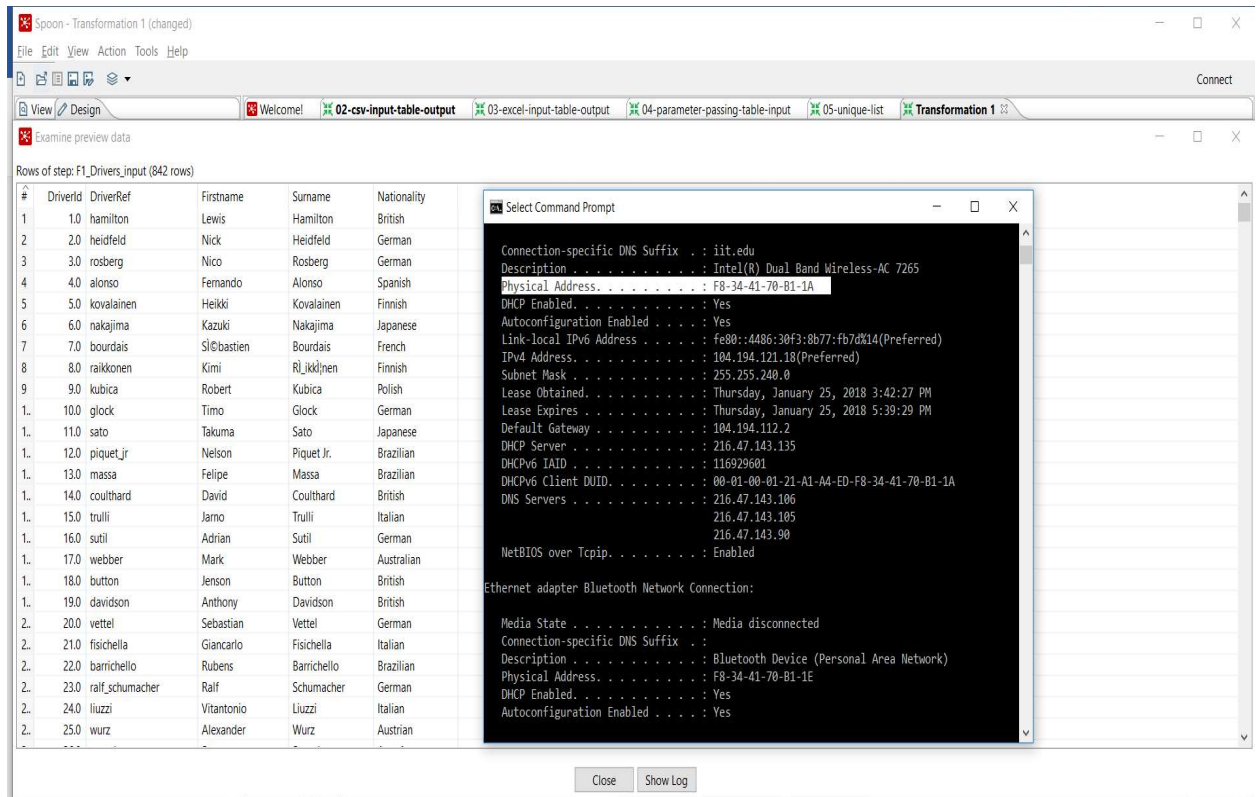
```
Autoconfiguration Enabled . . . . : Yes
Wireless LAN adapter Wi-Fi:
Connection-specific DNS Suffix  . : iit.edu
Description . . . . . : Intel(R) Dual Band Wireless-AC 7265
Physical Address. . . . . : F8-34-41-70-B1-1A
DHCP Enabled. . . . . : Yes
Autoconfiguration Enabled . . . . : Yes
Link-local IPv6 Address . . . . . : fe80::4486:30f3:8b77:fb7d%14(Preferred)
IPv4 Address. . . . . : 104.194.121.18(Preferred)
Subnet Mask . . . . . : 255.255.240.0
Lease Obtained. . . . . : Thursday, January 25, 2018 3:42:27 PM
Lease Expires . . . . . : Thursday, January 25, 2018 5:39:29 PM
Default Gateway . . . . . : 104.194.112.2
DHCP Server . . . . . : 216.47.143.135
DHCPv6 IAID . . . . . : 116929601
DHCPv6 Client DUID. . . . . : 00-01-00-01-21-A1-A4-ED-F8-34-41-70-B1-1A
DNS Servers . . . . . : 216.47.143.106
                        216.47.143.105
                        216.47.143.90
NetBIOS over Tcpip. . . . . : Enabled
Ethernet adapter Bluetooth Network Connection:
```

2. Then in Pentaho tool I have drag and drop Microsoft excel as an input.

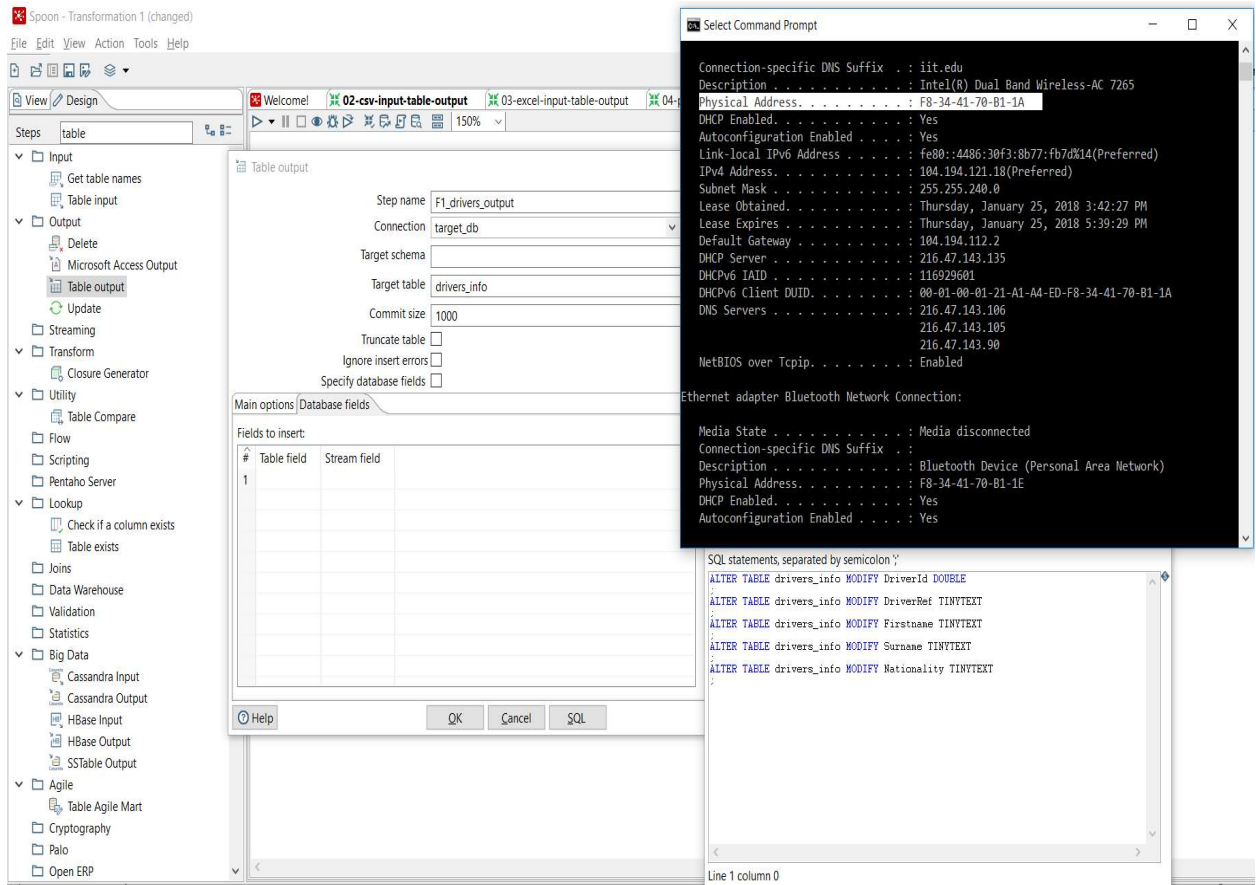


3. I have browse a xlsx file from the directory and uploaded in Excel input.

4. Preview of 1000 rows from the excel.

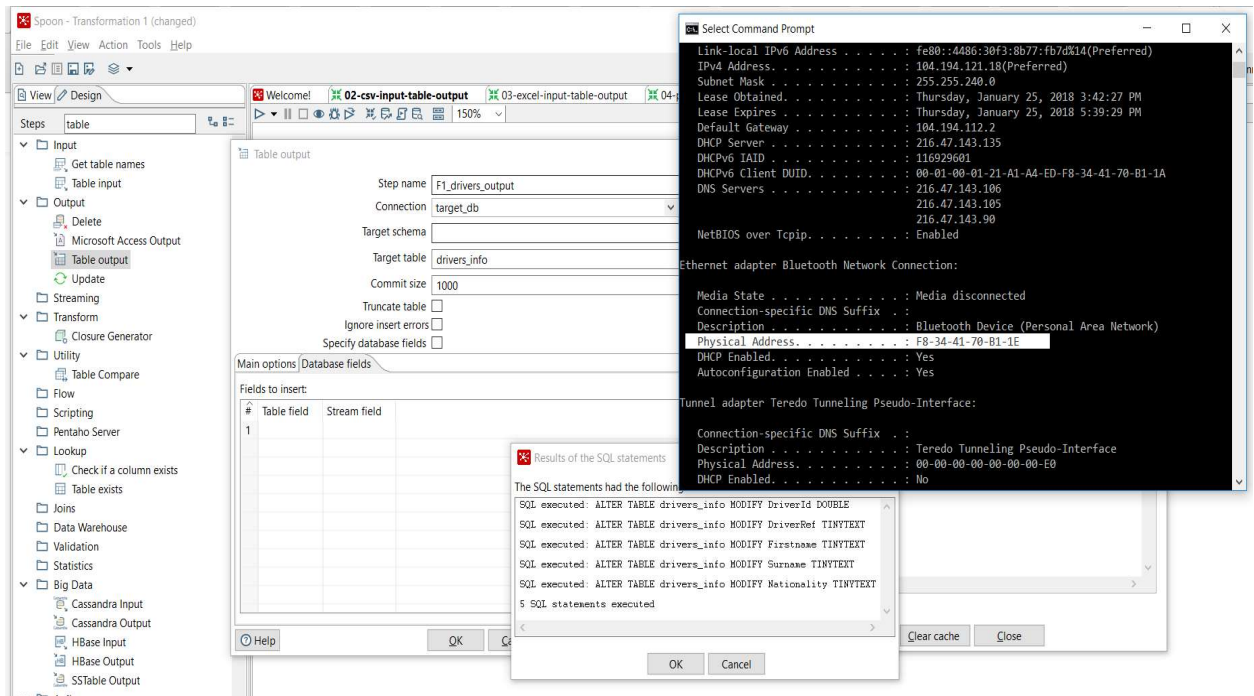


5. Selected Spread Sheet type(engine) as Excel 2007 XLSX (Apache POI), so that pentaho should know the type of input file.
6. In the field tab we can get the header row of xlsx input file.
7. Now I have taken table output in the transformation.
8. I have named it as “F1_drivers_output” connected with my DB “target_db” and target_table as “drivers_info”.
9. Now I moved on to other tab Database fields where I need to perform mapping of fields.

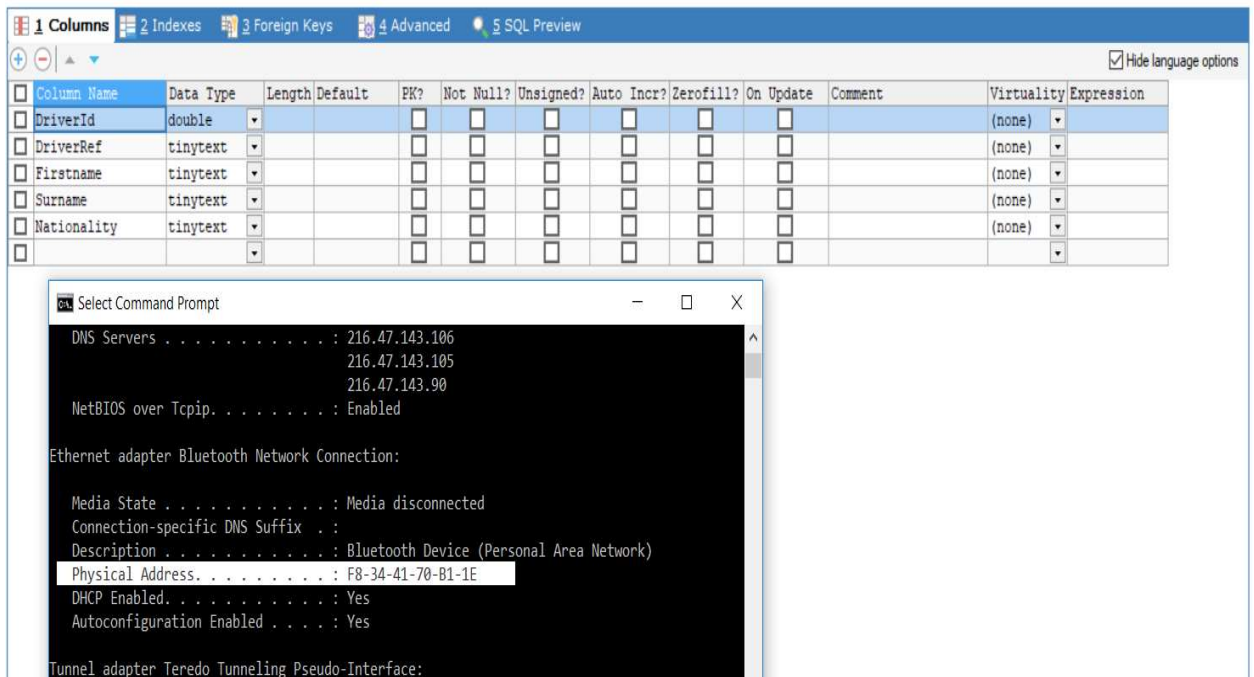


10. Now, I clicked on the button SQL for mapping of fields as we can see from the screenshot Pentaho changes the datatype as there was extra spaces defined so it converted it into double and Tinytext.

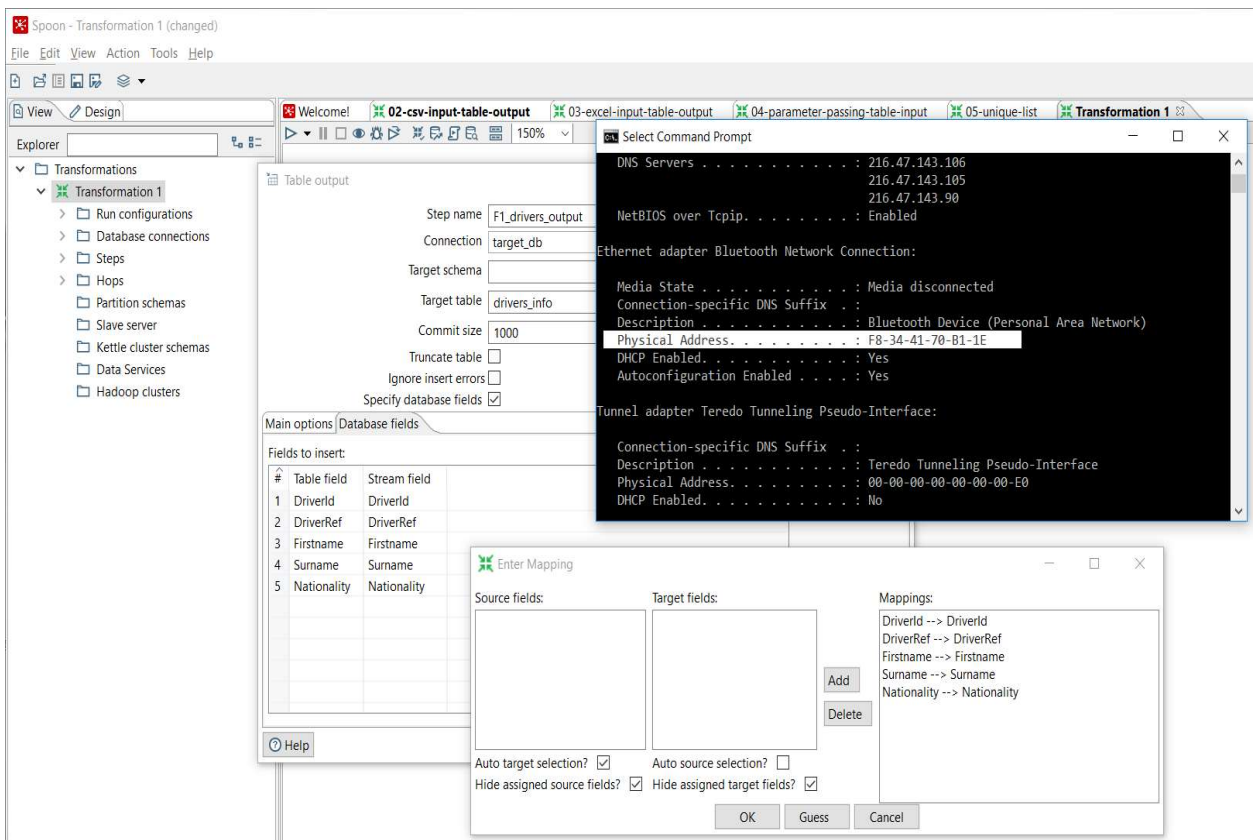
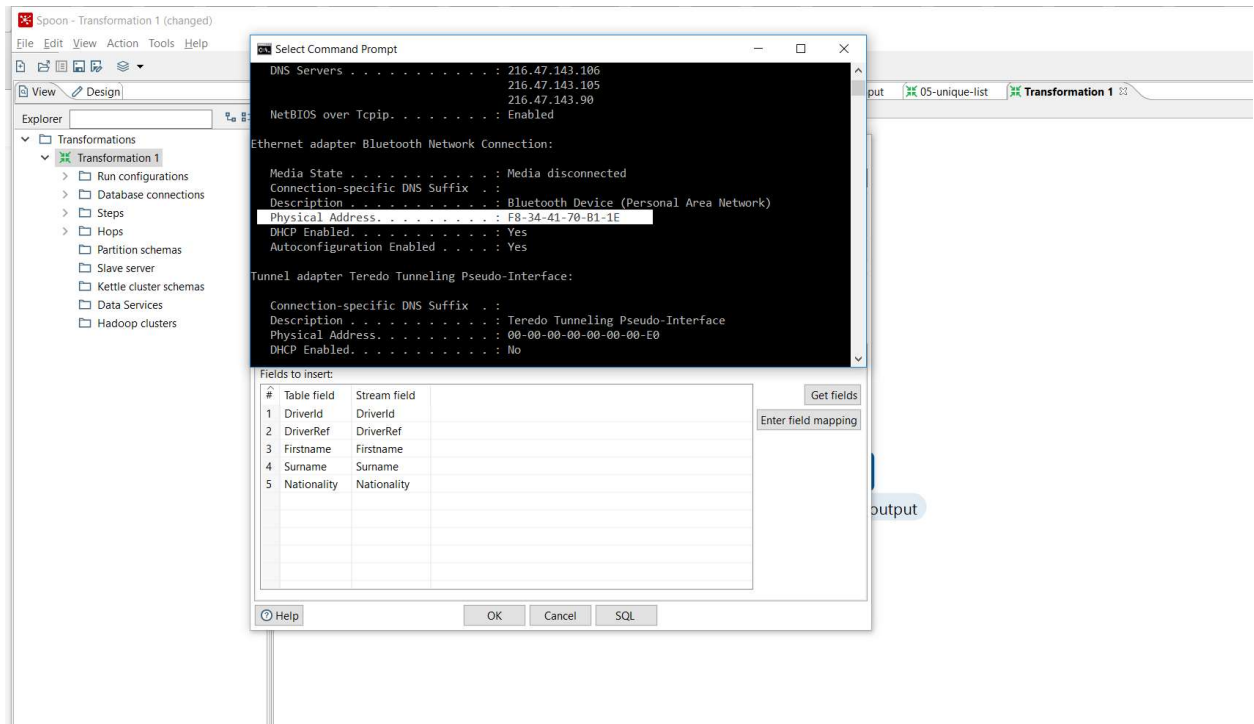
11. Thus, all the SQL statements are executed successfully by Pentaho.



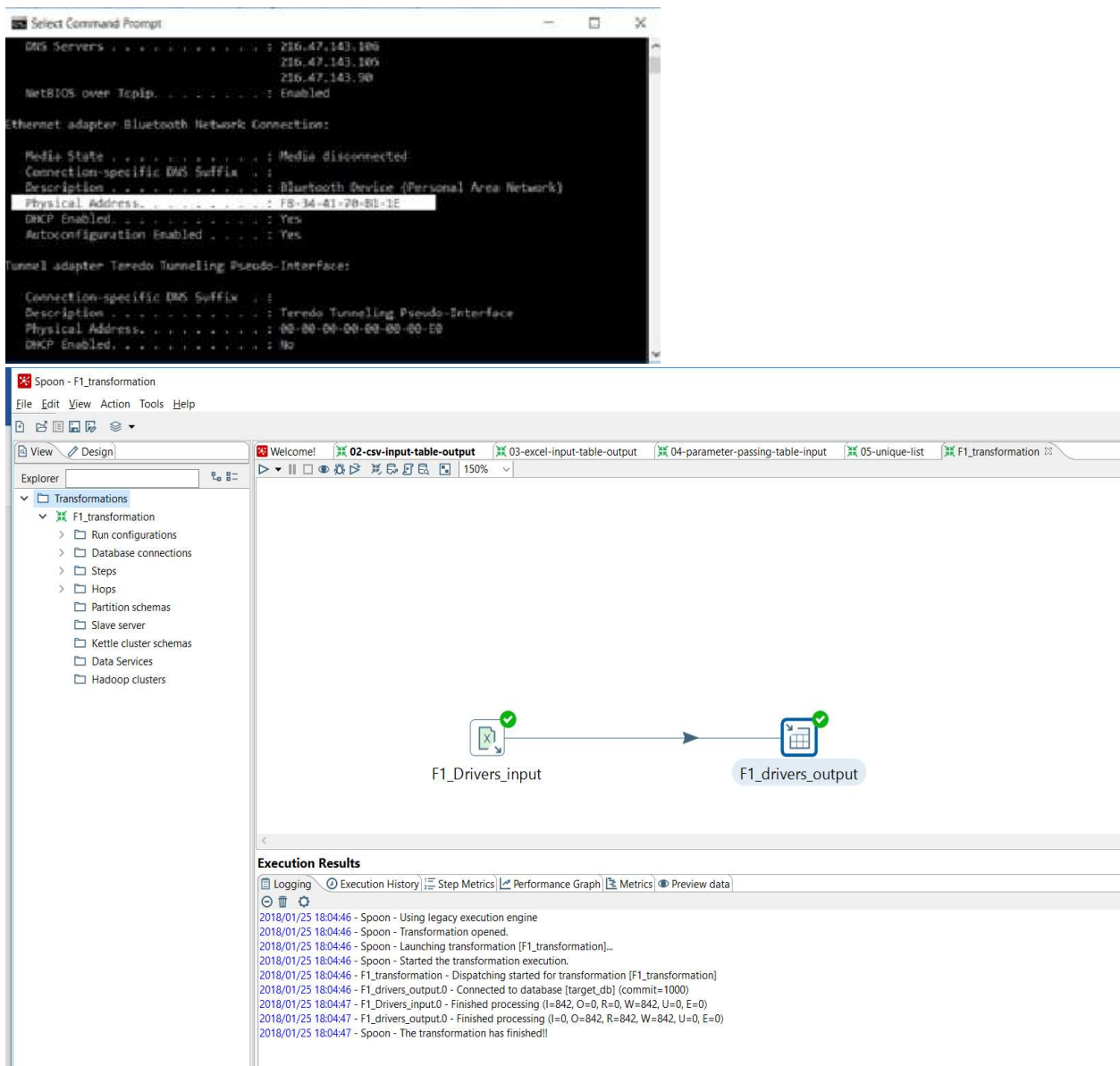
12. We can see the changes done in SQLYog.



13. Mapping is done in this we can see this by below screenshot.



14. Once the mapping is done we can the run the transformation, below is the screenshot of successfully done transformation, in total 842 rows are read and write without any error.



15. We can check in MYSQL DB by selecting all rows of the table, hence we found all 842 rows are fetched by the below screenshot, hence the transformation is done successfully.


```

14
15 SELECT * FROM target_db.drivers_info;
16

```

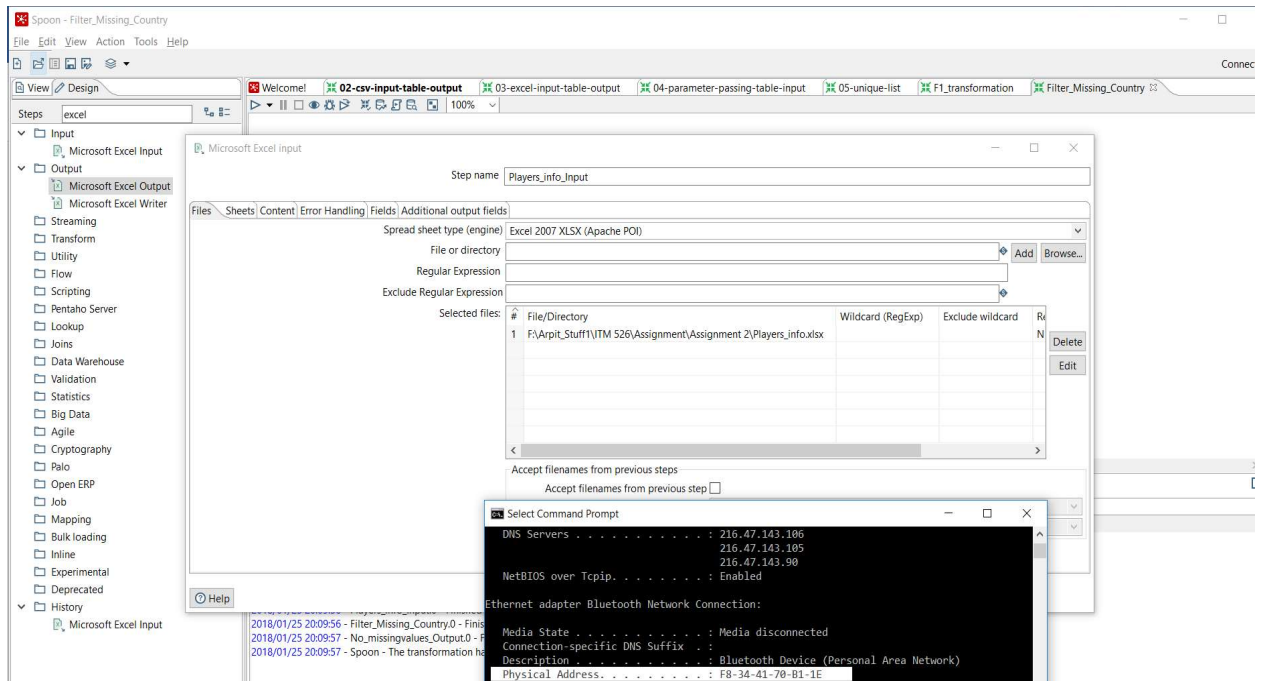
DriverId	DriverRef	Firstname	Surname	Nationality
1	hamilton	Lewis	Hamilton	British
2	heidfeld	Nick	Heidfeld	German
3	rosberg	Nico	Rosberg	German
4	alonso	Fernando	Alonso	Spanish
5	kovalainen	Heikki	Kovalainen	Finnish
6	nakajima	Kazuki	Nakajima	Japanese
7	bourdais	Sébastien	Bourdais	French
8	raikkonen	Kimi	Räikkönen	Finnish
9	kubica	Robert	Kubica	Polish
10	glock	Timo	Glock	German
11	sato	Takuma	Sato	Japanese
12	piquet_jr	Nelson	Piquet Jr.	Brazilian
13	massa	Felipe	Massa	Brazilian
14	coulthard	David	Coulthard	British
15	trulli	Jarno	Trulli	Italian
16	sutil	Adrian	Sutil	German
17	webber	Mark	Webber	Australian
18	button	Jenson	Button	British
19	davidson	Anthony	Davidson	British
20	vettel	Sebastian	Vettel	German
21	fisichella	Giancarlo	Fisichella	Italian
22	barrichello	Rubens	Barrichello	Brazilian
23	ralf_schum...	Ralf	Schumacher	German
24	liuzzi	Vitantonio	Liuzzi	Italian
25	wurz	Alexander	Wurz	Austrian

select * from target_db.drivers_info LIMIT 0, 1000

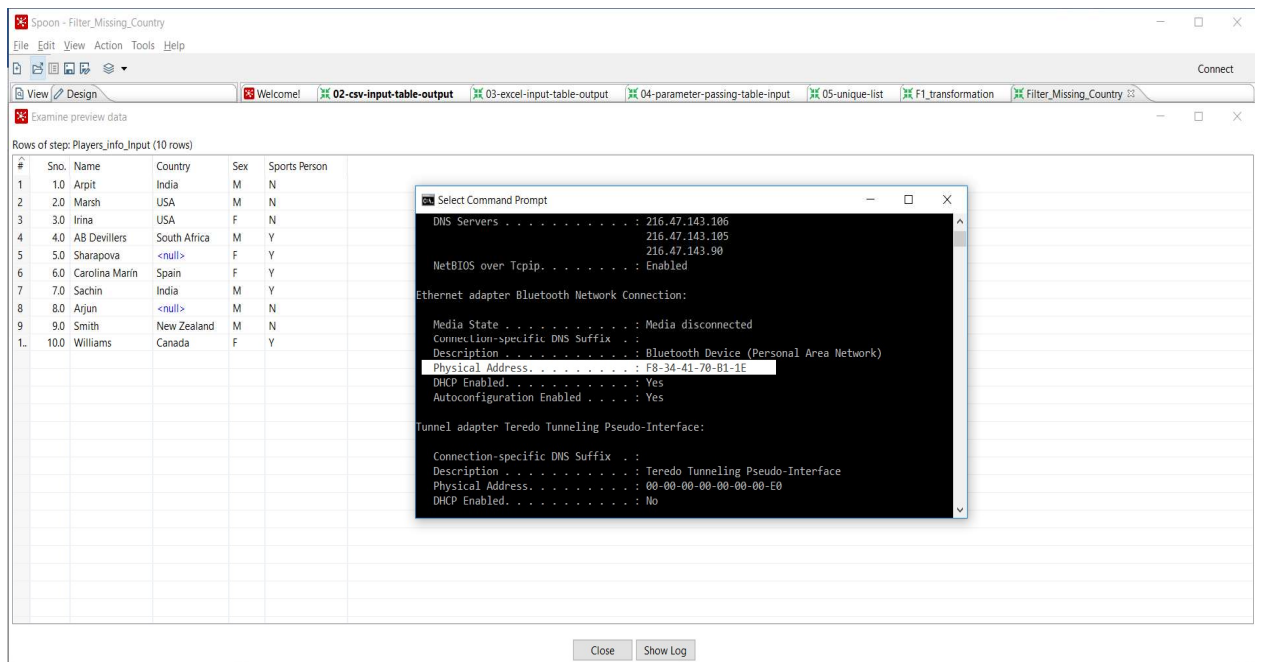
Exec: 0.003 sec Total: 0.006 sec 842 row(s) Connections: 1 [Upgr](#)

B) Filter rows with null values Transformation

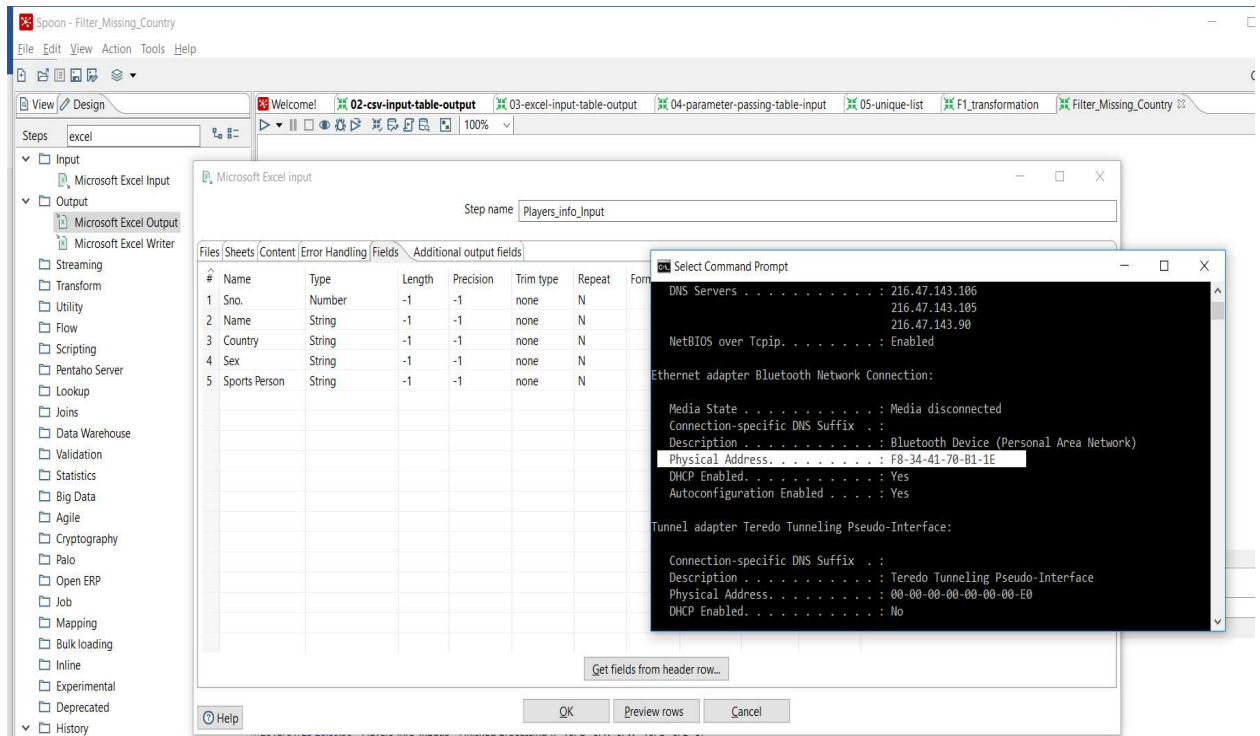
1. Firstly, we have taken input as Microsoft Excel as an input namely "Players_info_input" which includes some missing values for column Country.



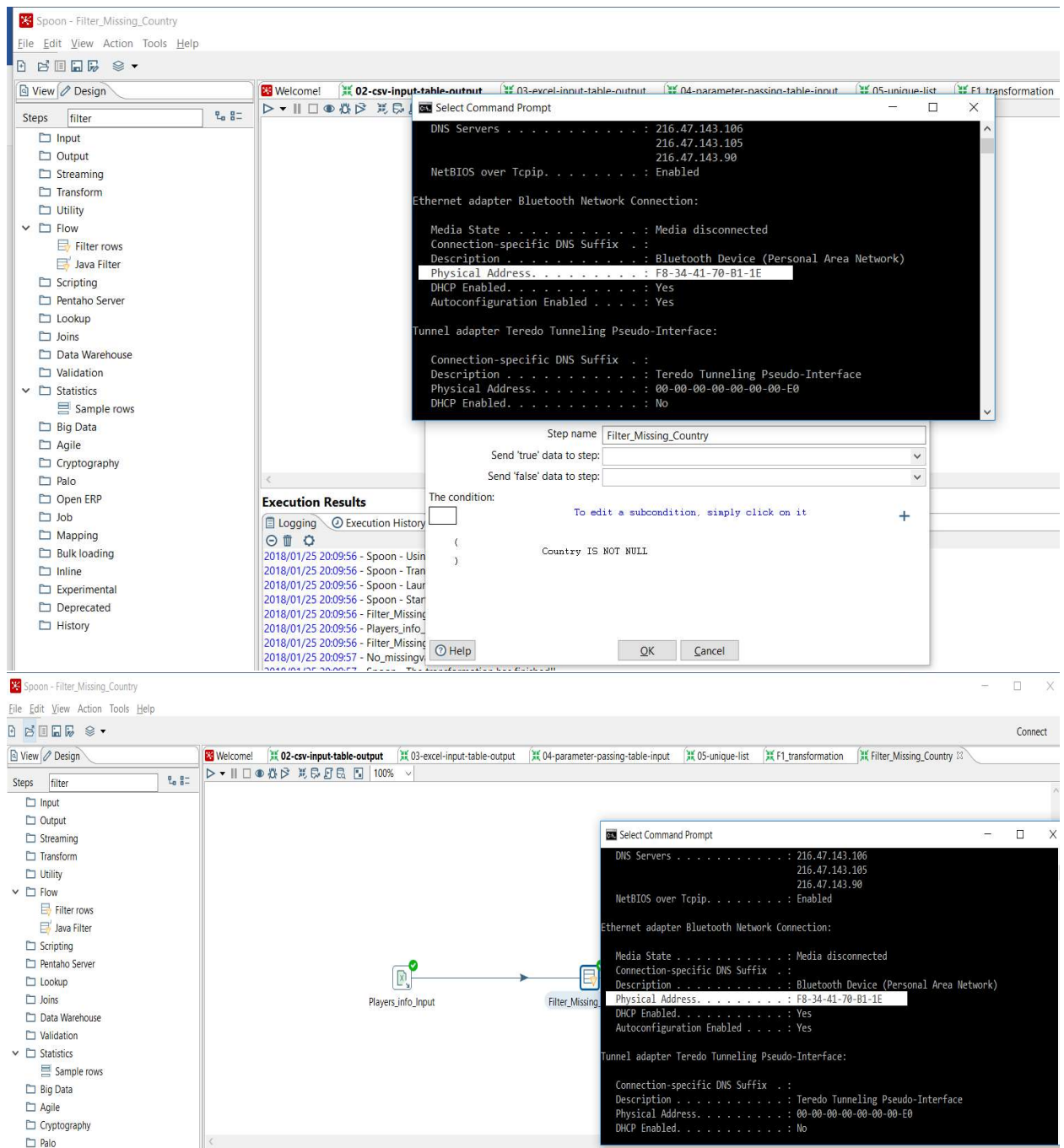
2. In the tab fields we get the header rows, below is the screenshot for this:



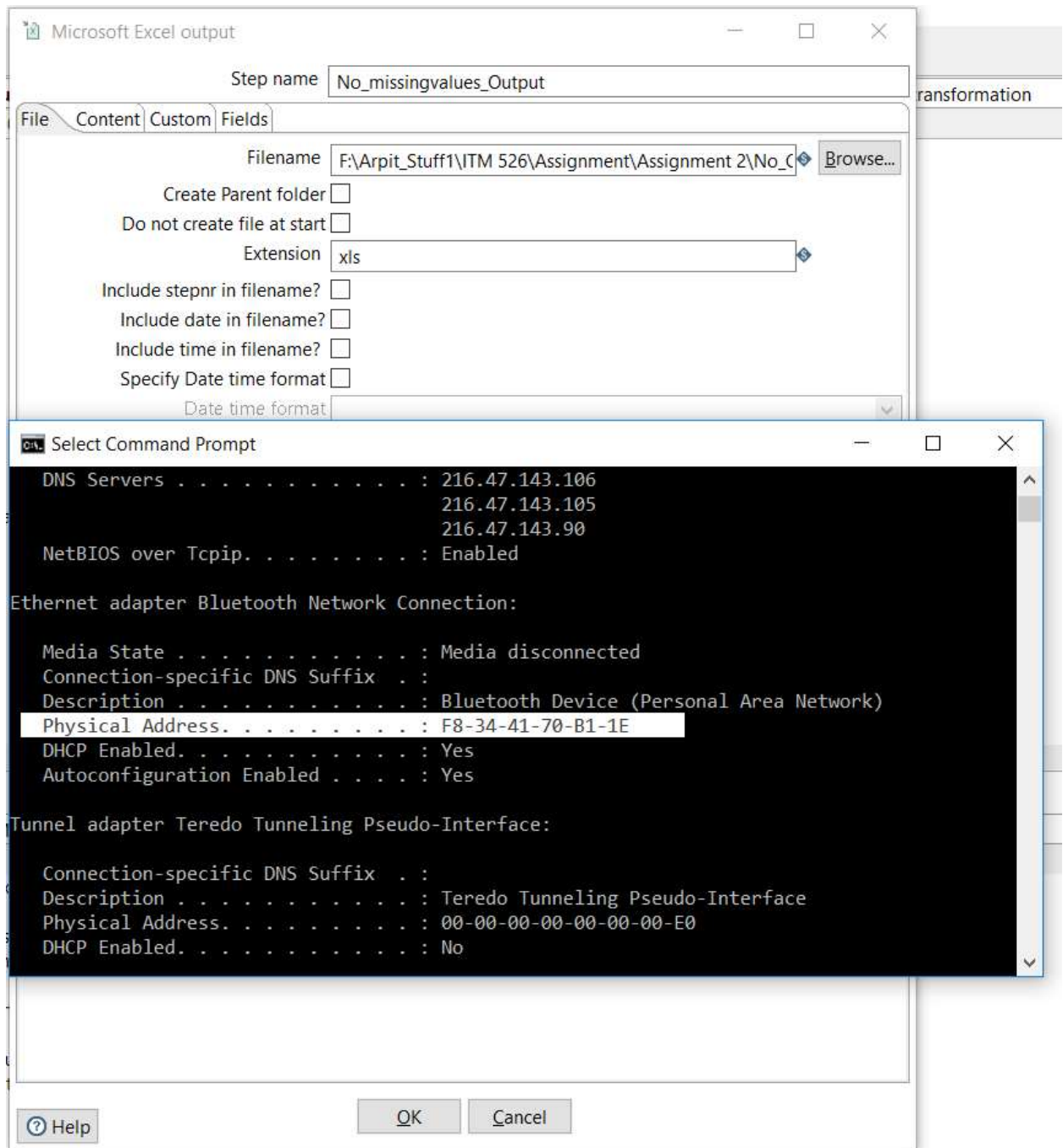
3. In the fields tab we get the data from the header rows below is the screenshot for this.



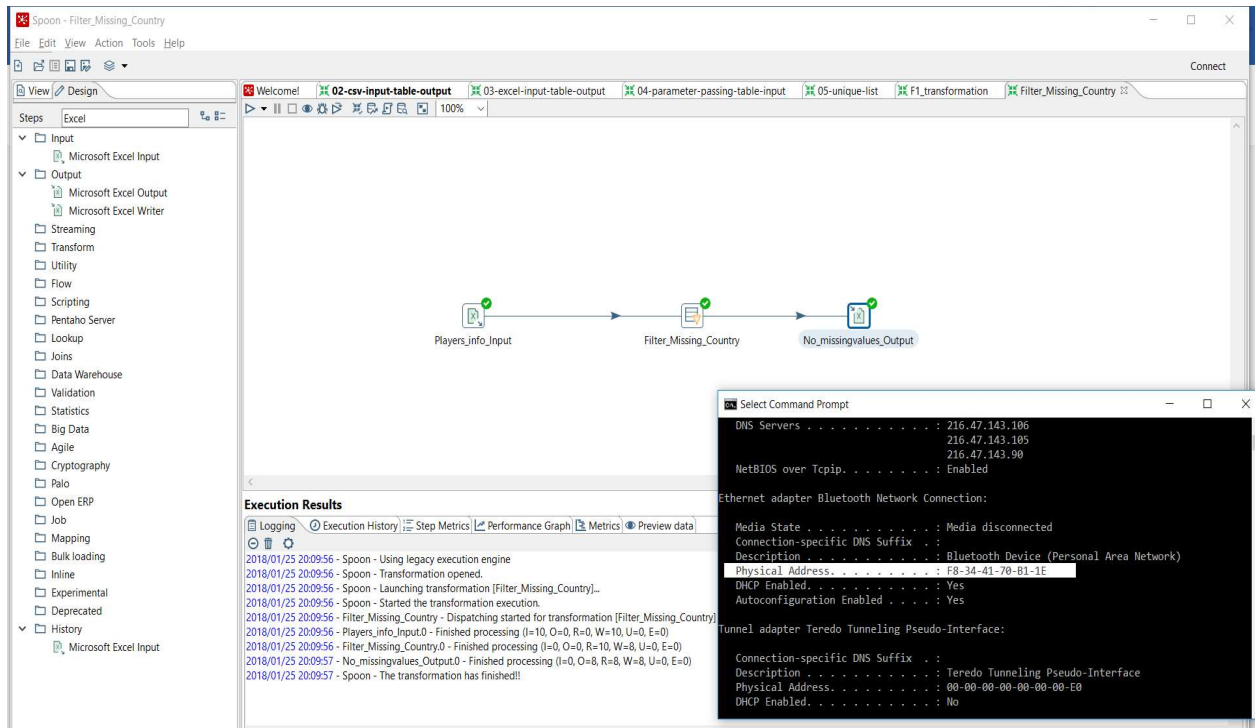
4. Now, adding filter rows to filter out the missing Country values from the excel sheet naming "Filter_Missing_Country".
5. Setting the condition in this step as "Country is not null", and then click on OK.



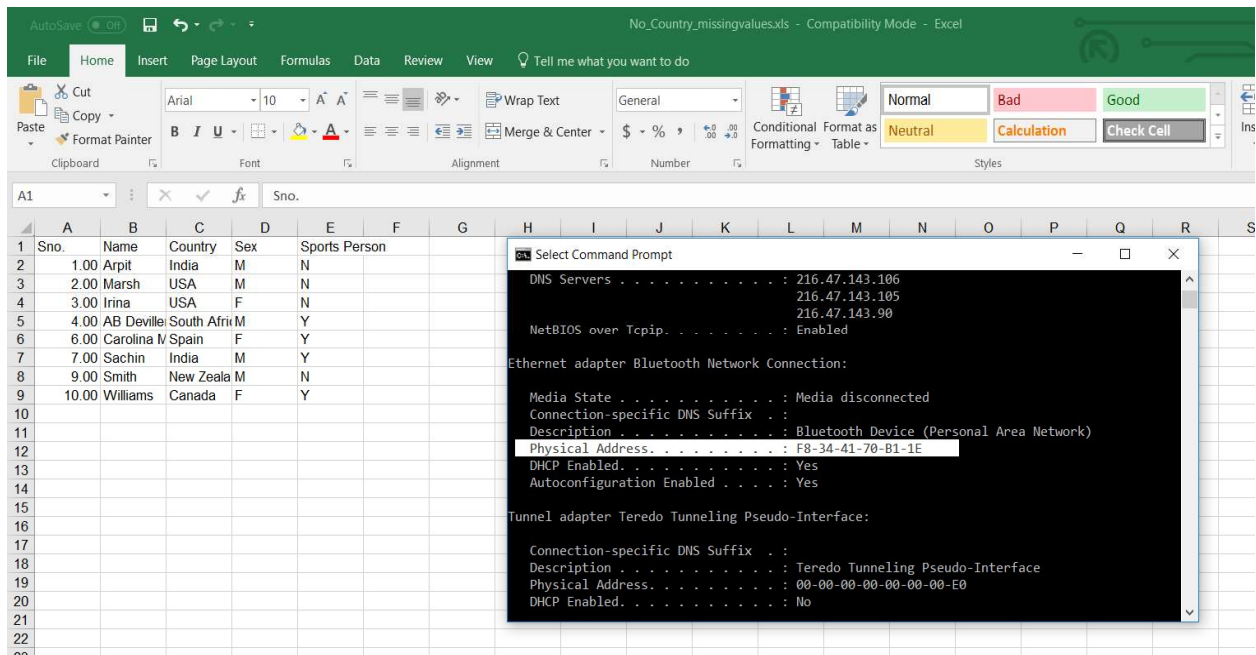
6. Now, taking all filtered values without any null values to an excel sheet output namely **"No_missingvalues_Output"**
7. Storing the excel in local workspace.



8. Now, transformation ran successfully without any errors.



9. Checking output excel file “No_missingvalues_Output”, there should be no null values in this excel sheet. Below is the screenshot for this.



10. Hence Transformation done successfully.