SSIS Assignment -1

Task 1: Integration with ETL Data Warehouse (DWH)

Scenario: Your company has a data warehouse designed to consolidate data from various sources for analytical purposes. You need to create an SSIS package that extracts data from a transactional database and loads it into the data warehouse.

Requirements:

- Create a Connection Manager to connect to the transactional database and the data warehouse.
- Extract Data from a transactional table using an OLE DB Source.
- Transform Data:
  Apply necessary transformations such as data type conversions, data cleansing, and calculations.
- Load Data into the data warehouse

OUTPUT

	IMERNAME	PHONE	ADDRESSLINE1	CITY	COUNTRY	TERRITORY	CONTACTLASTNAME	CONTACTFIRSTNAME	DEALSIZE
1	f Toys Inc.	2125557818	897 Long Airport Avenue	NYC	USA	NA	Yu	Kwai	Small
2	Collectables	26.47.1555	59 rue de l'Abbaye	Reims	France	EMEA	Henriot	Paul	Small
3	ouveniers	+33 1 46 62 7555	27 rue du Colonel Pierre Avia	Paris	France	EMEA	Da Cunha	Daniel	Medium
4	rownUps.com	6265557265	78934 Hillside Dr.	Pasadena	USA	NA	Young	Julie	Medium
5	ste Gift Ideas Co.	6505551386	7734 Strong St.	San Francisco	USA	NA	Brown	Julie	Medium
6	cs Stores Inc.	6505556809	9408 Furth Circle	Burlingame	USA	NA	Hirano	Juri	Medium
7	lus Designs Imports	20.16.1555	"184		59000	France	EMEA	Rance	Martine,Small
8	Gifts	+47 2267 3215	"Drammen 121		N 5804	Norway	EMEA	Oeztan	Veysel,Medium
9	heels Co.	6505555787	5557 North Pendale Street	San Francisco	USA	NA	Murphy	Julie	Small
10	anal Petit	(1) 47.55.6555	"25		75016	France	EMEA	Perrier	Dominique,Medium
11	ilian Collectors	Co."	03 9520 4555	Level 3	3004	Australia	APAC	Ferguson	Peter,Medium
12	ome Inc.	2125551500	2678 Kingston Rd.	NYC	USA	NA	Frick	Michael	Small
13	Collectables Inc.	2015559350	7476 Moss Rd.	Newark	USA	NA	Brown	William	Medium
14	pot Inc.	2035552570	25593 South Bay Ln.	Bridgewater	USA	NA	King	Julie	Medium
15	helle Gifts	40.67.8555	"67		44000	France	EMEA	Labrune	Janine,Medium

Task 2: Data Warehouse Migrations

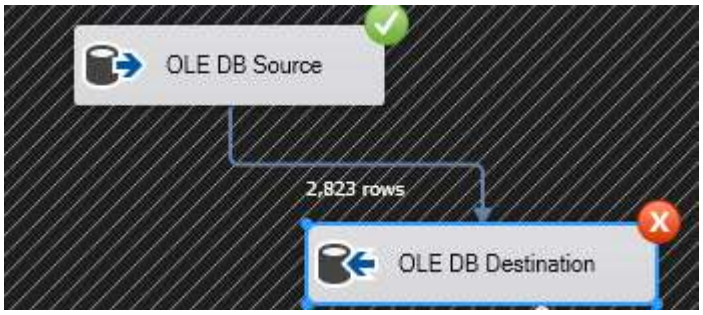
Scenario: Your organization is migrating its data warehouse from one server to another. You need to create an SSIS package that facilitates this migration.

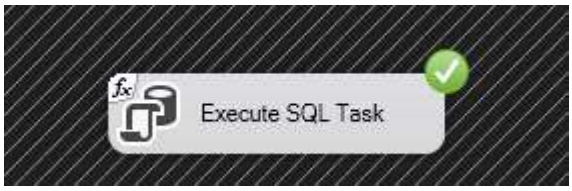
Requirements:

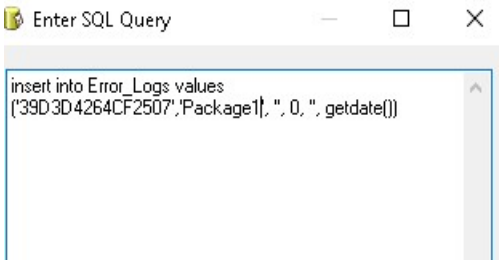
- Create Connection Managers for both the source and destination data warehouses.

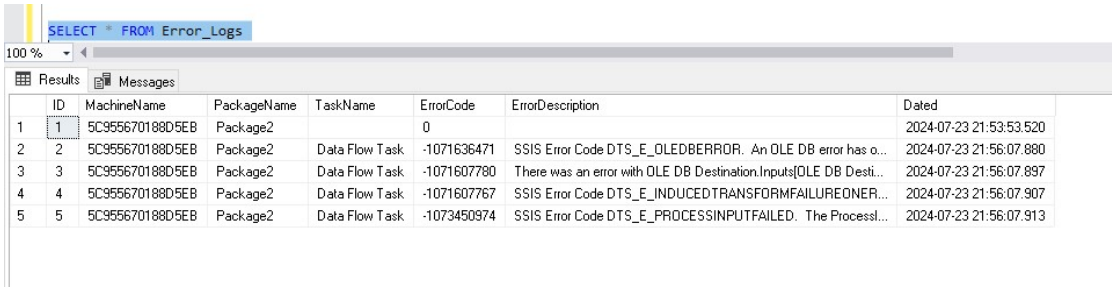
- Transfer Data from the source data warehouse to the destination using the Data Flow Task.
- Ensure Data Integrity:
    - Include checks and balances to ensure data is correctly migrated.
    - Log the success or failure of the migration process.

OUTPUT









	ID	MachineName	PackageName	TaskName	ErrorCode	ErrorDescription	Dated
1	1	5C955670188D5EB	Package2		0		2024-07-23 21:53:53.520
2	2	5C955670188D5EB	Package2	Data Flow Task	-1071636471	SSIS Error Code DTS_E_OLEDBERROR. An OLE DB error has o...	2024-07-23 21:56:07.880
3	3	5C955670188D5EB	Package2	Data Flow Task	-1071607780	There was an error with OLE DB Destination.Inputs[OLE DB Desti...	2024-07-23 21:56:07.897
4	4	5C955670188D5EB	Package2	Data Flow Task	-1071607767	SSIS Error Code DTS_E_INDUCEDTRANSFORMFAILUREONER...	2024-07-23 21:56:07.907
5	5	5C955670188D5EB	Package2	Data Flow Task	-1073450974	SSIS Error Code DTS_E_PROCESSINPUTFAILED. The Processl...	2024-07-23 21:56:07.913

Task 3: Implementing a Pivot Transformation

Scenario: You have data in a normalized format and need to pivot it for reporting purposes.

Requirements:

- Extract Data from the source table using an OLE DB Source.
- Apply a Pivot Transformation to transform the normalized data into a pivoted format.
- Load the Pivoted Data into a destination table.

OUTPUT

	YEAR_ID	jan	oct	nov	dec	feb	march	april	may	jun	july	aug	sept
1	2004	3245	NULL	NULL	NULL	3061	1978	2077	2618	2971	3174	4564	3171
2	2003	1357	5515	10179	2489	1449	1755	1993	2017	1649	1725	1974	2510
3	2005	3395	NULL	NULL	NULL	3393	3852	2634	4357	NULL	NULL	NULL	NULL
4	2004	NULL	5483	10678	3804	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL

Task 4: Incremental Load

Scenario: To optimize ETL processes, you need to implement an incremental load to update only the changed data in the data warehouse.

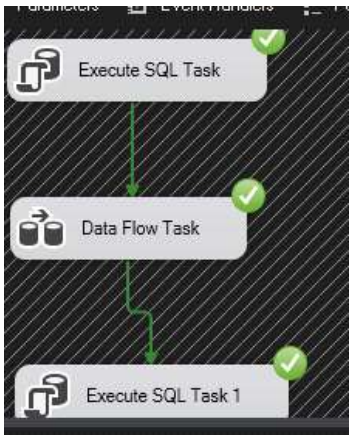
Requirements:

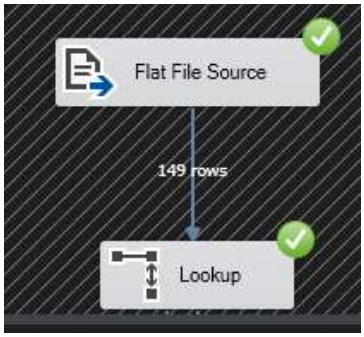
- Identify Changed Data: Use methods such as timestamps, change data capture using lookup, or checksums.
- Extract Only the Changed Data from the source.
- Update the Data Warehouse with the new and changed data only.

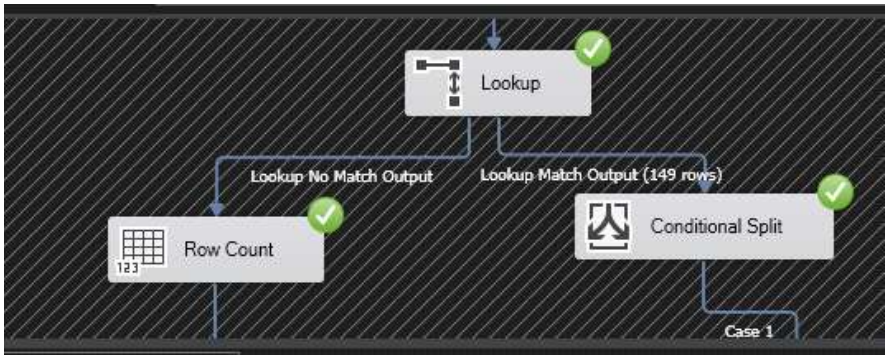
QUERY

```
IF NOT EXISTS(Select * from sys.objects WHERE object_id =
OBJECT_ID(N'[dbo].[audit_log_table]') AND type in (N'U'))
CREATE table audit_log_table(Id int identity, PackageName varchar(200),TableName
varchar(200),RecordsInserted INT, RecordsUpdated INT, DATED Datetime);
GO
IF NOT EXISTS(Select * from sys.objecacts WHERE object_id =
OBJECT_ID(N'[dbo].[Inputtable]') AND type in (N'U'))
CREATE TABLE [dbo].[Inputtable](
	[airport_id] [int] NULL,
	[name] [varchar](50) NULL,
	[city] [varchar](50) NULL,
	[state] [varchar](50) NULL
) ON [PRIMARY]
GO
IF EXISTS(Select * from sys.objects WHERE object_id =
OBJECT_ID(N'[dbo].[Updatedtable]') AND type in (N'U'))
DROP TABLE [dbo].[Updatedtable]
GO
CREATE TABLE [dbo].[Updatedtable](
	[airport_id] [int] NULL,
	[name] [varchar](50) NULL,
	[city] [varchar](50) NULL,
	[state] [varchar](50) NULL
) ON [PRIMARY]
GO
```

OUTPUT







	Id	PackageName	TableName	RecordsInserted	RecordsUpdated	DATED
1	1	task4.dtsx	Inputtable	0	0	2024-07-26 21:15:43.157
2	2	task4.dtsx	Inputtable	1	0	2024-07-26 21:16:59.810

Results | Messages

	airport_id	name	city	state
1	10165	Adak	Adak Island	AK
2	10299	Ted Stevens Anchorage International	Anchorage	AK
3	10304	Aniak Airport	Aniak	AK
4	10754	Wiley Post/Will Rogers Memorial	Barrow	AK
5	10551	Bethel Airport	Bethel	AK
6	10926	Merle K Mudhole Smith	Cordova	AK
7	14709	Deadhorse Airport	Deadhorse	AK
8	11336	Dillingham Airport	Dillingham	AK
9	11630	Fairbanks International	Fairbanks	AK
10	11997	Gustavus Airport	Gustavus	AK
11	12523	Juneau International	Juneau	AK
12	12819	Ketchikan International	Ketchikan	AK
13	10245	King Salmon Airport	King Salm...	AK
14	10170	Kodiak Airport	Kodiak	AK
15	13970	Ralph Wien Memorial	Kotzebue	AK

Task 5: Transformations

Scenario: Your company needs to transform raw data into a format suitable for reporting. You need to perform multiple transformations within an SSIS package.

Requirements:

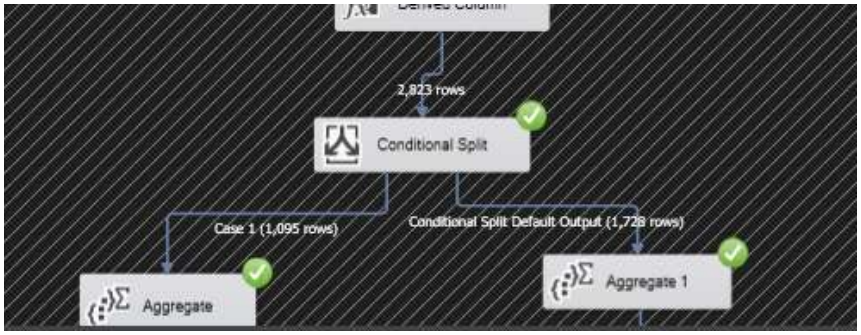
- Extract Data from a source table using an OLE DB Source.
- Apply Transformations such as:
  - Data Conversion
  - Derived Column
  - Conditional Split
  - Aggregate
- Load Transformed Data into a destination table.

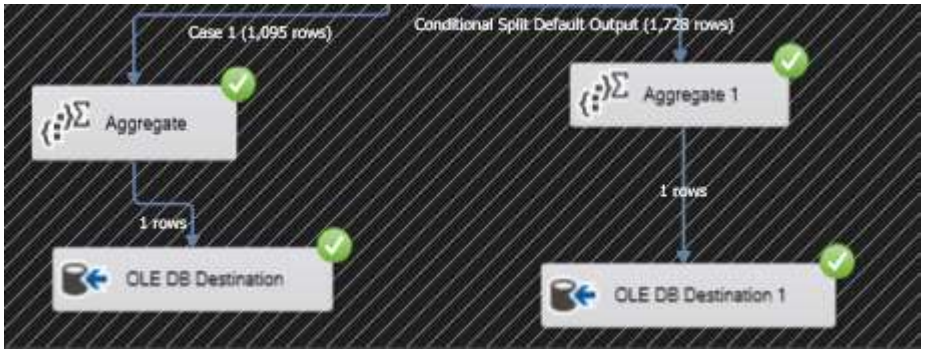
OUTPUT

Input Column	Output Alias	Data Type	Length	Precision	Scale	Code Page
QUANTITYORDERED	new_quantityordered	numeric [DT_NUMERIC]		18	0	
PRICEEACH	new_priceeach	numeric [DT_NUMERIC]		18	0	
MONTH_ID	new_monthid	numeric [DT_NUMERIC]		18	0	

Derived Column Name	Derived Column	Expression	Data Type	Le
Total	<add as new column>	new_quantityordered * new_priceeach	numeric [DT_NUMERIC]	

Input Column	Output Alias	Operation
Total	Total	Sum



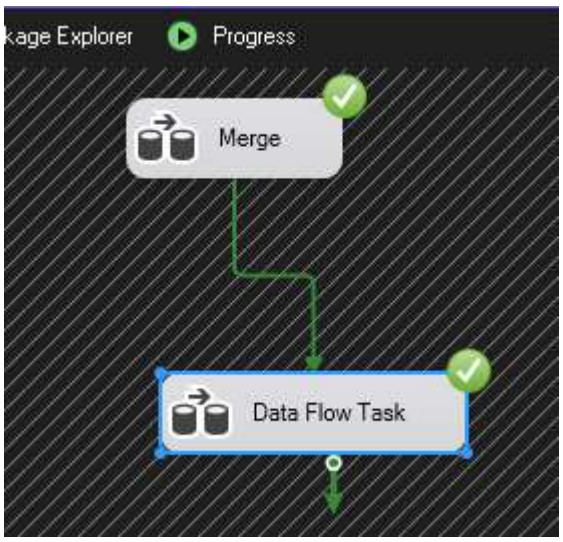


Task 6: MERGE & FUZZY LOOKUP
Scenario: You need to merge two datasets and use fuzzy matching to handle potential duplicates.

Requirements:

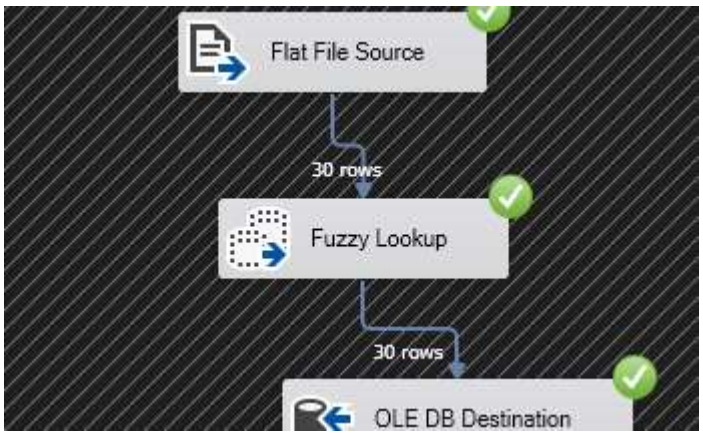
- Extract Data from two source tables using OLE DB Sources.
- Apply a Merge Join to combine the datasets based on a common key.
- Use Fuzzy Lookup to identify and resolve duplicates in the merged data.
- Load the Cleaned Data into a destination table.

OUTPUT


kage Explorer
Progress
Merge
Data Flow Task


OLE DB Source
15 rows
Sort
OLE DB Source 1
15 rows
Sort 1


15 rows
Merge
30 rows
OLE DB Destination 1



100 %

Results | Messages

	id	first_name	last_name	salary	jobdept	company	id (1)	_Similarity	_Confidence	_Similarity_id	_Similarity_first_name	_Similarity_last_name	_Simila
1	1	Ursuline	Rape	90437	Research and Development	Tr@upe	1	0.9570747	0.9307108	1	1	1	1
2	2	Yale	Dowles	78470	Legal	Meeveo	2	1	1	1	1	1	1
3	3	Lorrie	Peskett	61535	Business Development	Rhy.cero	3	0.9636405	0.9696509	1	1	1	1
4	4	Rubetta	McPeake	81369	Marketing	Jaxbean	4	1	1	1	1	1	1
5	5	Jaime	Burniston	96886	Engineering	Wik%izz	5	0.956962	0.9344336	1	1	1	1
6	6	Alix	Eades	81293	Business Development	Lazz	6	1	1	1	1	1	1
7	7	Casper	Casbolt	24960	Business Development	Viva	7	1	1	1	1	1	1
8	8	Giovanni	Ipwell	49270	Engineering	Bubblebox	8	1	1	1	1	1	1
9	9	Lottie	Slade	87132	Support	Voonyx	9	1	1	1	1	1	1
10	10	Dulcine	Crowcher	92358	Marketing	Roodel	10	1	1	1	1	1	1
11	11	Guss	Stailey	34839	Human Resources	M@eeveo	11	0.9645533	0.8473246	1	1	1	1
12	12	April	Grievson	64703	Marketing	Blogspan	12	1	1	1	1	1	1
13	13	Mabel	Hardern	76473	Research and Development	Ph##oto...	13	0.9742448	0.9780918	1	1	1	1
14	14	Marsiella	Cicero	87101	Services	Voonix	14	1	1	1	1	1	1
15	15	Theresina	Zannelli	96224	Accounting	Giga$zo...	15	0.9809481	0.9398407	1	1	1	1
16	16	Audrie	Brazur	27318	Business Development	Linklinks	16	1	1	1	1	1	1
17	17	Maribeth	Crewdson	81510	Engineering	Tanf$eed	17	0.9623417	0.9254951	1	1	1	1

Task 7: Using Script Task

Scenario: You need to perform a complex data transformation that is not supported by the standard SSIS components. A Script Task can be used to achieve this.

Requirements:

- Add a Script Task to the Control Flow.
- Write a Script: that performs the required transformation. e.g. Reading data from a file, processing it, and writing the results to a database table.
- Execute the Script Task within an SSIS package.

QUERY

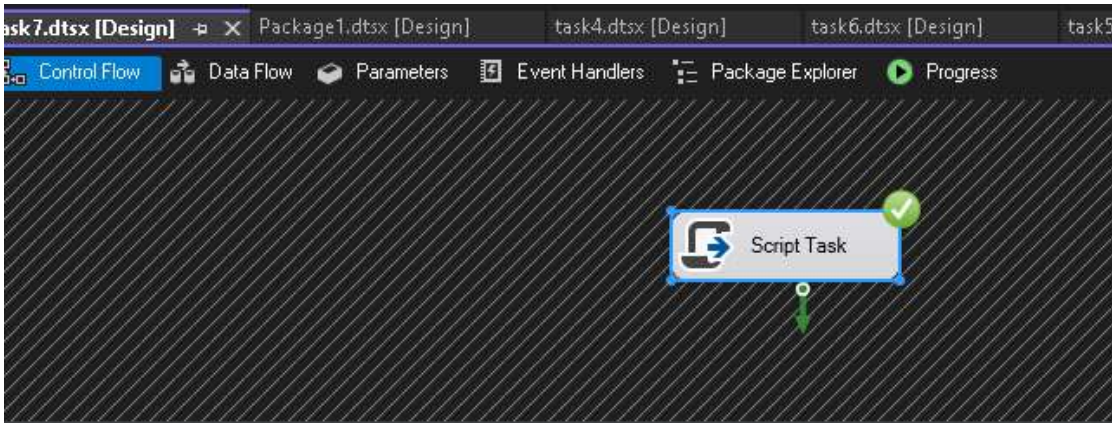
```
public void Main()
{
    try
    {
        // Set variables
        string serverName = "39d3d4264cf2507"; // Replace with your SQL
```

```
Server instance name

        string databaseName = "ass"; // Replace with your database name

        string tableName = "Airports"; // Replace with your table name

        string username = "sa"; // Replace with your SQL Server username

        string password = "pass@word1"; // Replace with your SQL Server
password

        string connectionString = $"Data Source={serverName};Initial
Catalog={databaseName};User ID={username};Password={password};";



        // Create SqlConnection using connection string

        using (SqlConnection conn = new SqlConnection(connectionString))

        {

            conn.Open(); // Open the connection



            // Get all CSV files in the specified directory

            string filePath = @"D:\ssisnew"; // Update filepath accordingly

            string[] fileEntries = Directory.GetFiles(filePath, "*.csv");



            // Process each file found

            foreach (string fileName in fileEntries)

            {

                using (StreamReader SourceFile = new
StreamReader(fileName))

                {

                    string Line;

                    int ctr = 0;



                    while ((Line = SourceFile.ReadLine()) != null)

                    {

                        if (ctr != 0) // Skip header if present (assuming
first line is header)

                        {

                            Line = Line.Trim();
```

```
                            // Split the line by comma and trim each field
                            string[] values = Line.Split(',');

                            // Construct the SQL query (using parameterized
query for safety)
                            string query = $"INSERT INTO {tableName}
(airport_id, city, state, name) " +
                                           $"VALUES (@AirportId, @City,
@State, @Name)";

                            // Execute SQL query with parameters
                            using (SqlCommand SQLCommand = new
SqlCommand(query, conn))
                            {
SQLCommand.Parameters.AddWithValue("@AirportId", values[0].Trim());
                                SQLCommand.Parameters.AddWithValue("@City",
values[1].Trim());
SQLCommand.Parameters.AddWithValue("@State", values[2].Trim());
                                SQLCommand.Parameters.AddWithValue("@Name",
values[3].Trim());

                                SQLCommand.ExecuteNonQuery();
                            }
                        }
                        ctr++;
                    }
                }
            }

            // Set SSIS task result to success
            Dts.TaskResult = (int)ScriptResults.Success;
        }
    }
```

```
    catch (Exception ex)
    {
        // Log error and set SSIS task result to failure
        Dts.Events.FireError(0, "Exception from Script Task", ex.Message +
"\r" + ex.StackTrace, String.Empty, 0);
        Dts.TaskResult = (int)ScriptResults.Failure;
    }
}
```

OUTPUT



Results | Messages

	airport_id	city	state	name
1	10135	Allentown/Bethlehem/Easton	PA	Lehigh Valley International
2	10136	Abilene	TX	Abilene Regional
3	10140	Albuquerque	NM	Albuquerque International Sunport
4	10141	Aberdeen	SD	Aberdeen Regional
5	10146	Albany	GA	Southwest Georgia Regional
6	10154	Nantucket	MA	Nantucket Memorial
7	10155	Waco	TX	Waco Regional
8	10157	Arcata/Eureka	CA	Arcata
9	10158	Atlantic City	NJ	Atlantic City International
10	10165	Adak Island	AK	Adak
11	10170	Kodiak	AK	Kodiak Airport
12	10185	Alexandria	LA	Alexandria International
13	10208	Augusta	GA	Augusta Regional at Bush Field
14	10245	King Salmon	AK	King Salmon Airport
15	10257	Albany	NY	Albany International
16	10268	Waterloo	IA	Waterloo Regional
17	10279	Amarillo	TX	Rick Husband Amarillo Internati...
18	10299	Anchorage	AK	Ted Stevens Anchorage Interna...
19	10304	Aniak	AK	Aniak Airport
20	10329	Naples	FL	Naples Municipal
21	10361	Watertown	NY	Watertown International
22	10372	Aspen	CO	Aspen Pitkin County Sardy Field