

# LeadGen

DATA DRIVEN SALES  
INTELLIGENCE AT YOUR  
FINGERTIPS

---

(A TECHNICAL  
METHODOLOGY BRIEFING)

Maximum Bushel Estimate

# Our sales teams need to know:

---

What potential customers exist in the sales region?

Which are the most important to us?

Where are these customers located?

How can we get in touch with them?

LeadGen answers all of these questions –

- With an intuitive, data driven decision-making dashboard for our sales team
- With advanced predictive and geoanalytics algorithms on the back end

# LeadGen Dashboard

1. Easily digestible, sortable, list of leads in the sales region

- Features contact info, farm acres, and min and max bushel estimates given local grain silos

2. Map visualization indicates who and where are the best leads

3. Scatterplot representing possible anomalies in the dataset

For any dashboard element – click on a lead to explore that lead on the other elements

## LeadGen - Data Driven Sales Intelligence

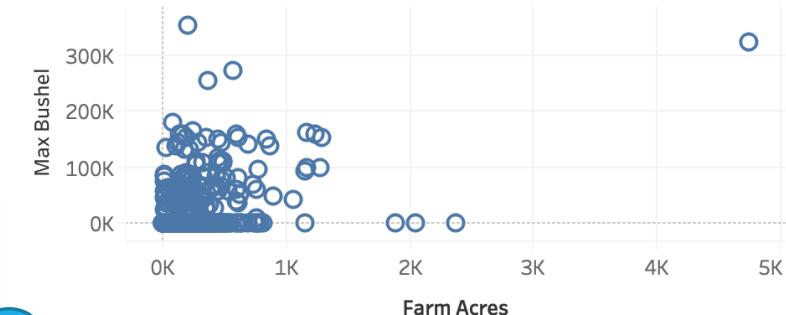
Scroll the list, explore the map, inspect possible outliers. Find the best leads.

We're proud to provide grain storage estimates (min & max bushels) sourced from aerial imagery, county parcel data, and predictive algorithms.

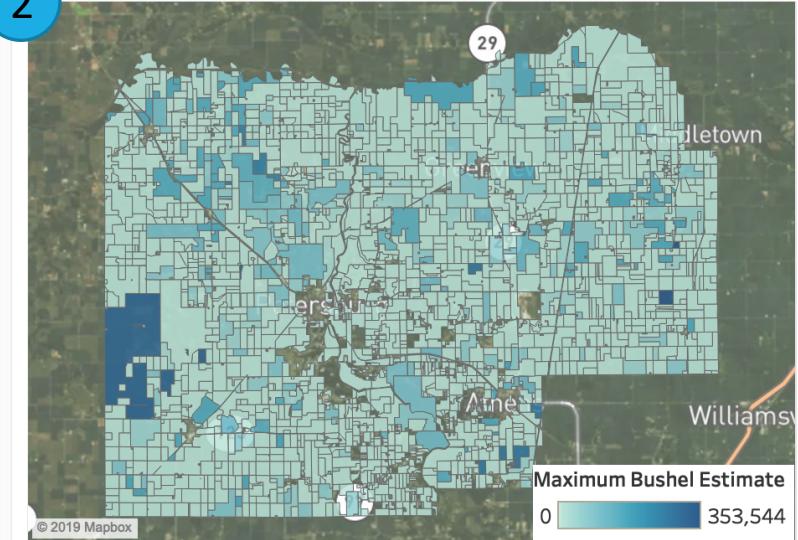
1	Contact Info	Max Bu..	Min Bushel	Farm Acres
WELL CHESTER L	16986 PEORIA ST, A...	353,544	160,009	195
GRIGSBY FAMILY PAR..	10736 MARKERT AV...	324,690	147,990	4,739
CARTER LESTER B & I..	24945 ETHELL AV, A...	272,480	119,008	565
WINKELMANN WAYN..	14944 WHITES CRO...	254,612	111,927	355
WEIDHUNER HOMER ..	20492 PEORIA ST, G...	179,599	76,327	74
DAVIS M PATRICK & P..	10901 REICHART AV...	163,516	69,131	238
TUCKER ANN P & RON..	1 PICKERING LN, SP...	160,431	61,330	1,161
DIGIOVANNA STEPHE..	13441 POST ST, ATH...	159,249	70,192	163
SWAAR ROBERT M & ..	25903 LEVEE ST, GR...	158,492	71,982	134
TODD JAMES H TRUS..	29953 SWEETWATE...	157,816	67,756	1,228
SMITH STEVEN A	502 N ELM ST, TALL...	157,433	70,107	594
THOMAS MICHAEL & ..	13632 WATKINS SC...	153,449	69,363	1,280
ATTERBERRY LYMAN ..	528 W JACKSON ST, ...	153,368	66,274	341
ALLEN STEPHEN C & S..	178 PEMBERTON LO...	152,126	66,098	190
BERGMAN STEVEN C ..	10131 COUNTY RD 3...	151,751	69,255	600
HAWKS BLANCHE E T..	113 APRICOT PT, PE...	151,041	63,316	837
TODD DONALD T & DO..	18836 ENGEL ST, AT...	150,364	66,661	446
DIGIOVANNA DANNY	22502 DIGOVANNA ..	144,785	65,250	112
KUNKEN DENNIS L & ..	29818 IRISH GROVE ..	144,750	60,088	463
WINKELMANN ROBE..	17556 CASSEN AV, ..	143,174	64,623	278
PAYDON WILLIAM DE..	26466 W 135TH ST ..	139,406	56,758	689

3

Outlier Detection interface - high storage, low farm acres may indicate anomalies



2



# LeadGen - Data Driven Sales Intelligence

Scroll the list, explore the map, inspect possible outliers. Find the best leads.

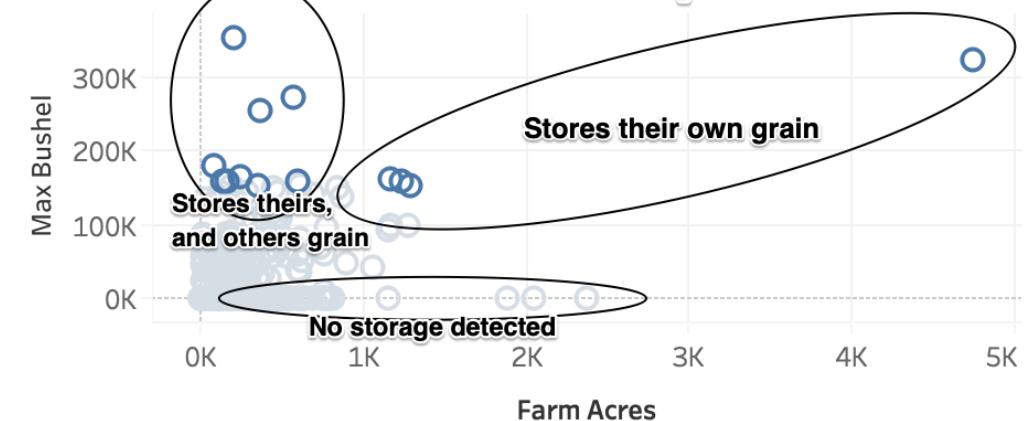
We're proud to provide grain storage estimates (min & max bushels) sourced from aerial imagery, county parcel data, and predictive algorithms.

**Filter for top leads**

Owner	Contact Info	Max Bu..	Min Bushel	Farm Acres
POWELL CHESTER L	16986 PEORIA ST, A..	353,544	160,009	195
GRIGSBY FAMILY PAR..	10736 MARKERT AV..	324,690	147,990	4,739
CARTER LESTER B & I..	24945 ETHELL AV, A..	272,480	119,008	565
WINKELMANN WAYN..	14944 WHITES CRO..	254,612	111,927	355
WEIDHUNER HOMER ..	20492 PEORIA ST, G..	179,599	76,327	74
DAVIS M PATRICK & P..	10901 REICHART AV..	163,516	69,131	238
TUCKER ANN P & RON..	1 PICKERING LN, SP..	160,431	61,330	1,161
DIGIOVANNA STEPHE..	13441 POST ST, ATH..	159,249	70,192	163
SWAAR ROBERT M & ..	25903 LEVEE ST, GR..	158,492	71,982	134
TODD JAMES H TRUS..	29953 SWEETWATE..	157,816	67,756	1,223
SMITH STEVEN A	502 N ELM ST, TALL..	157,433	70,107	594
THOMAS MICHAEL & ..	13632 WATKINS SC..	153,449	69,363	1,280
ATTERBERRY LYMAN ..	528 W JACKSON ST, ..	153,368	66,274	341
ALLEN STEPHEN C & S..	178 PEMBERTON LO..	152,126	66,098	190
BERGMAN STEVEN C ..	10131 COUNTY RD 3..	151,751	69,255	600
HAWKS BLANCHE E T..	113 APRICOT PT, PE..	151,041	63,316	837
TODD DONALD T & DO..	18836 ENGEL ST, AT..	150,364	66,661	446
DIGIOVANNA DANNY	22502 DIGOVANNA ..	144,785	65,250	112
KUNKEN DENNIS L & ..	29818 IRISH GROVE ..	144,750	60,088	463
WINKELMANN ROBE..	17556 CASSEN AV, ..	143,174	64,623	278
PAYDON WILLIAM DE..	26466 W 135TH ST ..	139,406	56,758	689

Outlier Detection interface - high storage, low farm acres may indicate anomalies

**Understand customer profile**



**Know where to go, and how to split up sales teams geographically**

# LeadGen backend algorithms & development methodology

We will cover:

- Inputs
- Methods/Output

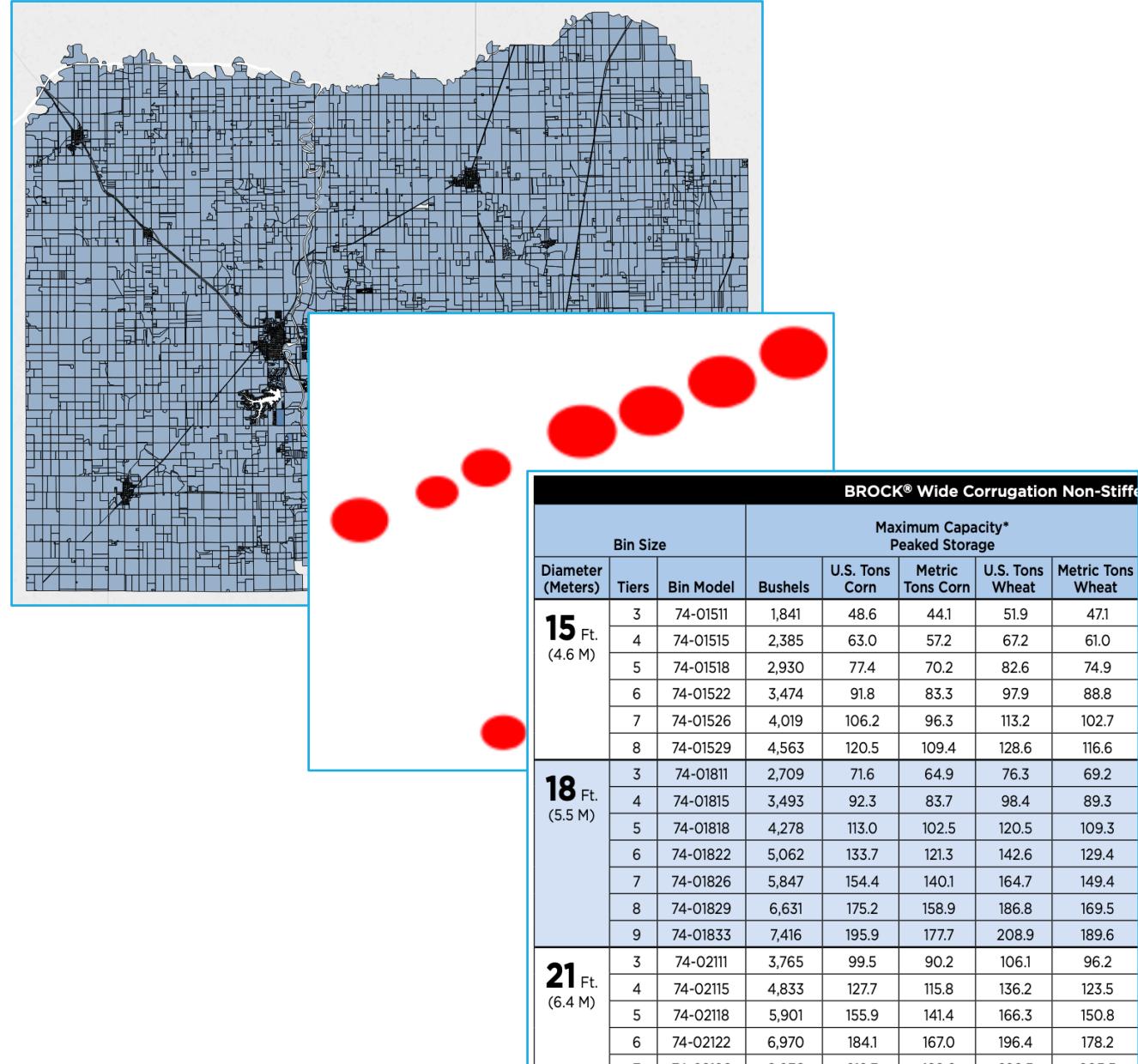
```
farm_acres_id_str = 'FARM_ACRES'
size_id_str = 'DIAMETER'
cluster_size = 3
cluster_distance_meters = 20
granger_silo_lookup_path = '../inputs/wide_corrugation_bin_data.csv'
output_path = '~/Desktop/lead_gen/test_output.shp'

#finds farm parcels, returns farm polygons and summed farm acreage grouped by owner
def return_farmer_gdf(parcel_shp_path, farmer_id_str, farm_acres_id_str):
    #load and clean parcel data and fields of interest
    parcel_df = gpd.read_file(parcel_shp_path)
    parcel_df[farmer_id_str] = parcel_df[farmer_id_str].str.strip()
    #assume NaN farm acres data is '0'
    parcel_df[farm_acres_id_str] = parcel_df[farm_acres_id_str].fillna(value=0)
    #subset parcel df for records with farm acres >0
    farm_parcels_df = parcel_df[parcel_df[farm_acres_id_str]>0]
    #subset df to fields of interest
    farm_parcels_by_owner_df = farm_parcels_df[[farmer_id_str,farm_acres_id_str,'geometry']]
    farm_parcels_contact_df = farm_parcels_df[[farmer_id_str,'MAILTO_ADD','MAILTO_CSZ']]
    farm_parcels_contact_df = farm_parcels_contact_df.set_index(farmer_id_str)
    farm_parcels_contact_df = farm_parcels_contact_df[~farm_parcels_contact_df.index.duplicated()]
    #group farm parcels by owner name, sum farm acres, dissolve geoms into single poly
    owner_farm_acres_and_geom = farm_parcels_by_owner_df.dissolve(by=farmer_id_str, aggfunc='sum')
    owner_farm_acres_and_geom.sort_values(by=[farm_acres_id_str], inplace=True, ascending=False)
    owner_farm_acres_and_geom = owner_farm_acres_and_geom.join(farm_parcels_contact_df)
    #print(owner_farm_acres_and_geom.columns)
    return owner_farm_acres_and_geom

#filters silo shapes for those with the correct diameter (4-20), and within a group of at least 3 (within 40 m
def return_silo_gdf(silo_shp_path, size_id_str, cluster_size, cluster_distance_meters):
    silo_df = gpd.read_file(silo_shp_path)
    silo_df['unique_id'] = silo_df.index
    #4 to 20 meters in diameter, 3 or more
    silo_size_subset_df = silo_df[(silo_df[size_id_str]>=4) & (silo_df[size_id_str]<=20)]
    #create 'reasonable distance' buffer around each potential silo - will use to determine proximity
    silo_subset_proj = silo_size_subset_df.copy()
    #remember to project data to meters!
    silo_subset_proj['geometry'] = silo_subset_proj['geometry'].to_crs(epsg=32616)
    silo_subset_proj['geometry'] = silo_subset_proj.geometry.buffer(cluster_distance_meters)
    #spatial match, exclude matches to self, group by, those with >2 matches are candidates
    silo_silo_spatial_matches = gpd.sjoin(silo_subset_proj, silo_subset_proj)
    silo_silo_spatial_matches = silo_silo_spatial_matches[['unique_id_left', 'unique_id_right']]
    count_per_silo = silo_silo_spatial_matches.groupby(['unique_id_left']).size()
    #filters for silos in clusters of at least 'cluster size argument'
    silo_id_series = count_per_silo[count_per_silo>=cluster_size]
    silo_id_list = silo_id_series.index.tolist()
    #filters original shapefile by ID list of detected silos
    screened_silo_df = silo_df[silo_df['unique_id'].isin(silo_id_list)]
```

# Inputs

1. County parcel dataset, with farm acreage, owner, and owner contact info attributes
2. Orthographic imagery derived silo footprints, with diameter estimates
3. Common silo dimension and capacity metrics (sourced from Granger)



# Methods (1)

## Explore and clean data

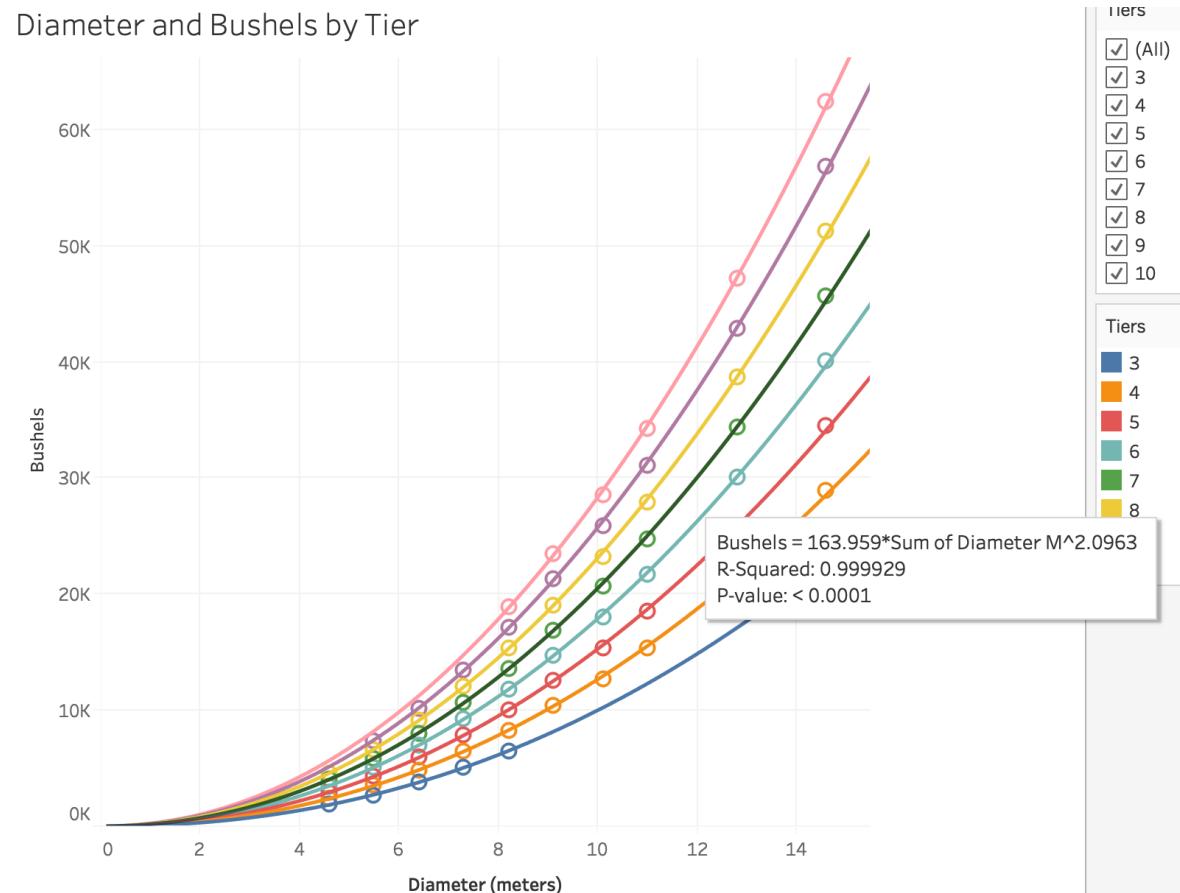
- 1. Examine/clean parcel fields**
  - a. Identified valuable fields related to owner, owner contact info, parcel farm acres, and spatial geometry
  - b. Replace Null values – assumed null for ‘farm acres’ was zero
  - c. Clean owner names – remove leading and trailing whitespace
  - d. Group by farmer name, sum acres, combine geometry
  - e. Function ‘return\_farmer\_gdf’ in ‘lead\_gen.py’
- 2. Examine/clean silo data**
  - a. Identified geometry and diameter (meters) as fields of interest
  - b. Removed silos that did not meet the expected diameter of 4-20 meters
  - c. Removed silos that did not fall within a max distance of 40m from at least two other silos (silos generally exist in clusters of 3 or more). Converted to meters based projection for distance.
  - d. Function ‘return\_silo\_gdf’ in ‘lead\_gen.py’
- 3. Source/construct sample silo metrics table from web**
  - a. Used all ‘Wide Corrugation’ bins data

## Methods (2)

Produce trend lines to estimate min and max bushel estimates per silo

**Granger data exhibited a clear trend when comparing silo diameter to silo bushel capacity:**

- This trend applies for all product tiers of Granger silos
- Higher tiers apply to larger diameters, and represent taller silos and more grain storage



## Methods (2)

Produce trend lines to estimate min and max bushel estimates per silo

**Next - these curves were created in Python and applied to our grain silos to produce min & max bushel estimates**

1. Performed per Granger tier
  - a. Gathered all data points per Granger tier
  - b. Converted all values with natural log to produce linear trended inputs
  - c. Trained a linear regression using these linear inputs to produce slope and intercept
  - d. Convert intercept value back into format for curve ( $e^x$ )
  - e. Estimated bushels: **bushels = intercept \* (diameter<sup>slope</sup>)**
2. For each silo, the nearest granger diameter was identified. This value was used to find the minimum and maximum tier available for that diameter
3. The curves for the min and max tiers for that diameter were then used with the actual diameter to identify a minimum and maximum bushels estimate

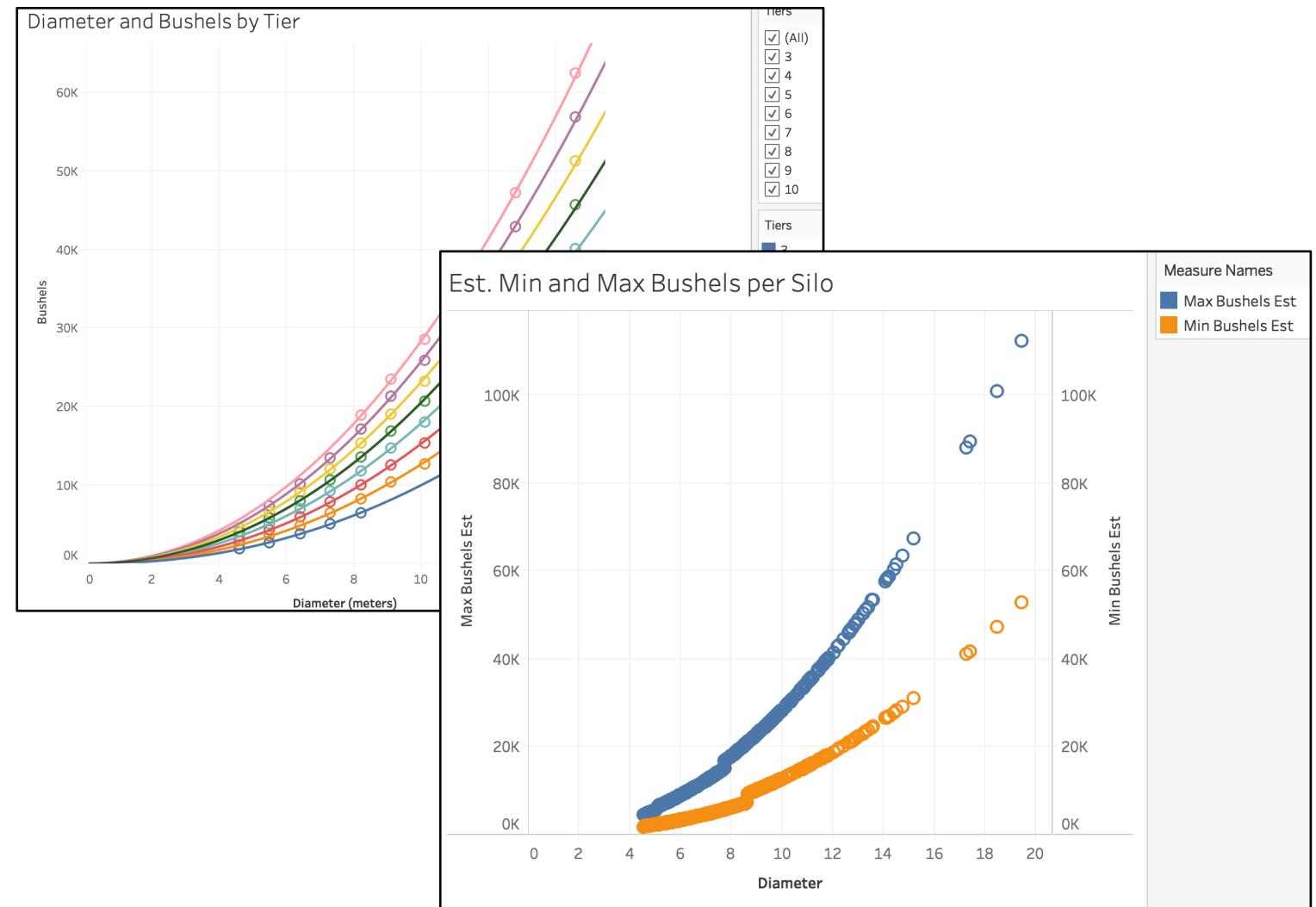
Functions used:

- ‘diameter\_bushel\_curve\_by\_granger\_tier’ in ‘lead\_gen.py’
- ‘min\_max\_granger\_tier\_by\_diameter’ in ‘lead\_gen.py’
- ‘silo\_min\_max\_bushel\_est’ in ‘lead\_gen.py’

## Methods (2)

Produce trend lines to estimate min and max bushel estimates per silo

Comparison of output 'silo bushel estimates' to original Granger curve



# Methods (3) & Output

## Matching silos to farmers & producing final leads dataset

### Silos were then matched to farmers

1. Our dataset of farmers (created in step 1) features their parcel footprint geometry. This geometry was matched with silos using a spatial join resulting in a list of silos (and their bushel capacity) per farmer
2. A groupby function then summed all silo capacity estimates for each farmer
3. Lastly, a farmer level dataset featuring ‘Owner’, ‘Contact Info’, ‘Max Bushels’, ‘Min Bushels’, and ‘Farm Acres’ is output as a shapefile
4. This file can be loaded, in seconds, into our Tableau Dashboard front end – empowering our sales team with a dynamic, easily updateable, lightweight tool

#	Abc test_output.shp <b>Farm Acres</b>	Abc test_output.shp <b>MAILTO_ADD</b>	=Abc Calculation <b>Contact Info</b>	Abc test_output.shp <b>Mailto Csv</b>	#	#	Abc test_output.shp <b>Min Bushel</b>	#	Abc test_output.shp <b>Max Bushel</b>	Abc test_output.shp <b>Owner</b>	Geome
	4,738.60	10736 MARKERT AV	10736 MARKERT AV, ...	TALLULA IL 62688	147,989.59		324,690.25		GRIGSBY FAMILY PA...	MULTIP	
	2,362.88	15702 SEASIDE LN	15702 SEASIDE LN, H...	HOUSTON TX 77062	0.00		0.00		AG HOLDINGS PLUS L...	POLYG	
	2,038.40	1213 E PINE ST	1213 E PINE ST, MAS...	MASON CITY IL 6266...	0.00		0.00		LOZIER JOHN G	MULTIP	
	1,874.83	10785 STATE HWY 78	10785 STATE HWY 7...	KEWANEE IL 61443	0.00		0.00		KINCAID MARGARET ...	MULTIP	
	1,279.62	13632 WATKINS SCH...	13632 WATKINS SCH...	OAKFORD IL 62673	69,363.19		153,448.72		THOMAS MICHAEL & ...	MULTIP	
	1,271.91	17762 SCHIRDING AV	17762 SCHIRDING AV...	PETERSBURG IL 626...	40,040.37		99,696.77		SCHIRDING FARMS INC	MULTIP	
	1,227.54	29953 SWEETWATER...	29953 SWEETWATER...	GREENVIEW IL 62642	67,756.16		157,815.58		TODD JAMES H TRUS...	MULTIP	
	1,161.19	1 PICKERING LN	1 PICKERING LN, SPR...	SPRINGFIELD IL 62712	61,330.02		160,431.34		TUCKER ANN P & RO...	MULTIP	
	1,150.61	17760 SCHIRDING AV	17760 SCHIRDING AV	PETERSONVILLE, ILLINOIS	25,000.00		20,700.00		SCHIRDING JOURNAL	MULTIP	

Function used: ‘match\_silo\_estimates\_to\_farmers’ in ‘lead\_gen.py’



Thank you for  
your interest in  
LeadGen.

Any questions?

---

Maximum Bushel Estimate