Preparing Your Dataset

Paweł Kordek
SOFTWARE ENGINEER

@pawel_kordek https://kordek.github.io

Dataset

What's inside

Loading

Preparing

Port of Embarkation	Sex	Cabin	Survived
Cherbourg	M	103	No
Queenstown	F	300	Yes
Cherbourg	F	245	No

No temporal data

Date	USD/EUR Exchange Rate
24.03.2017	0.81
25.03.2017	0.83
26.03.2017	0.82

Only one dimension



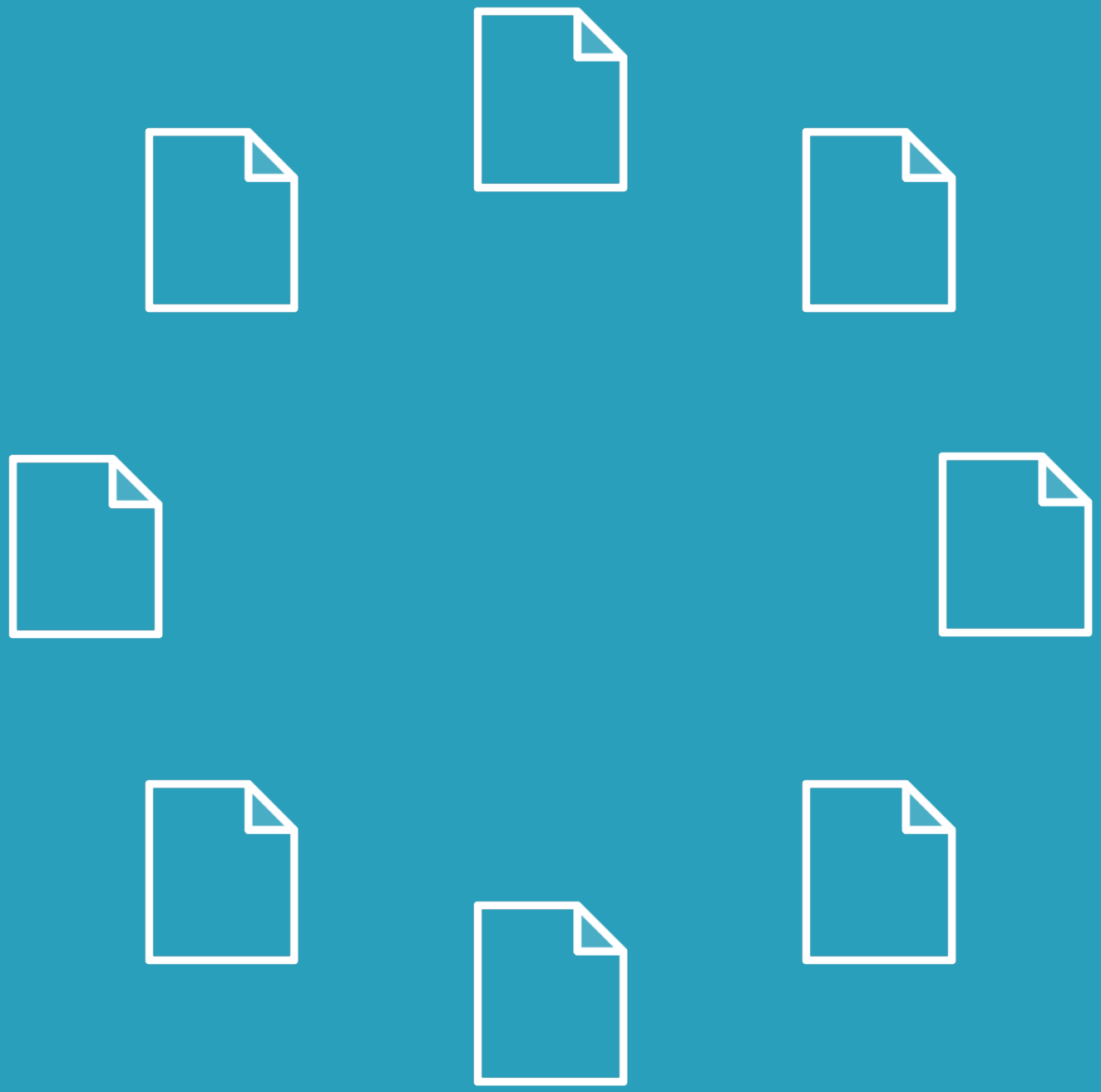
www.opensourcesports.com

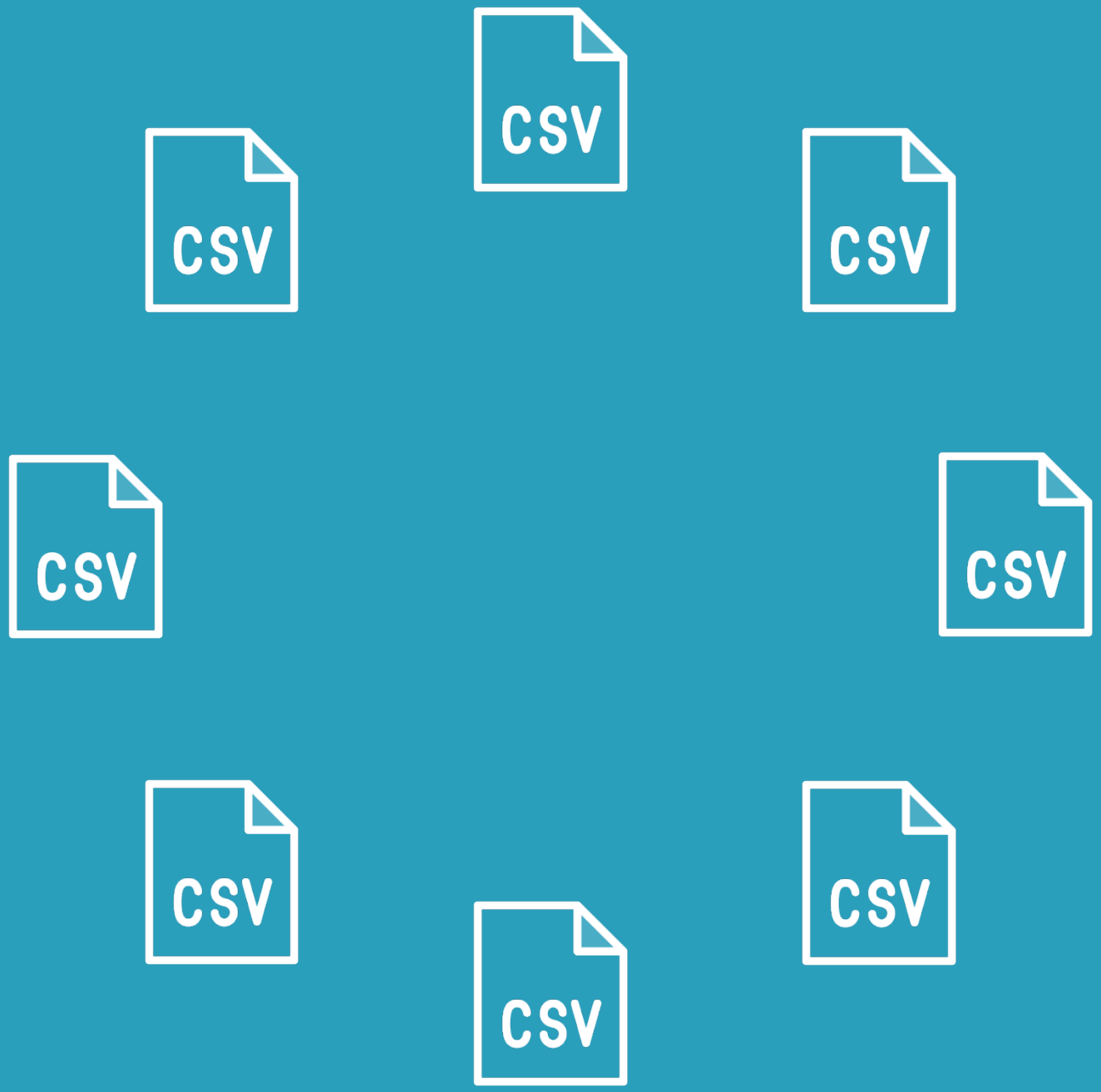
The information used herein was obtained free of charge from and is copyrighted by the Hockey Databank project.

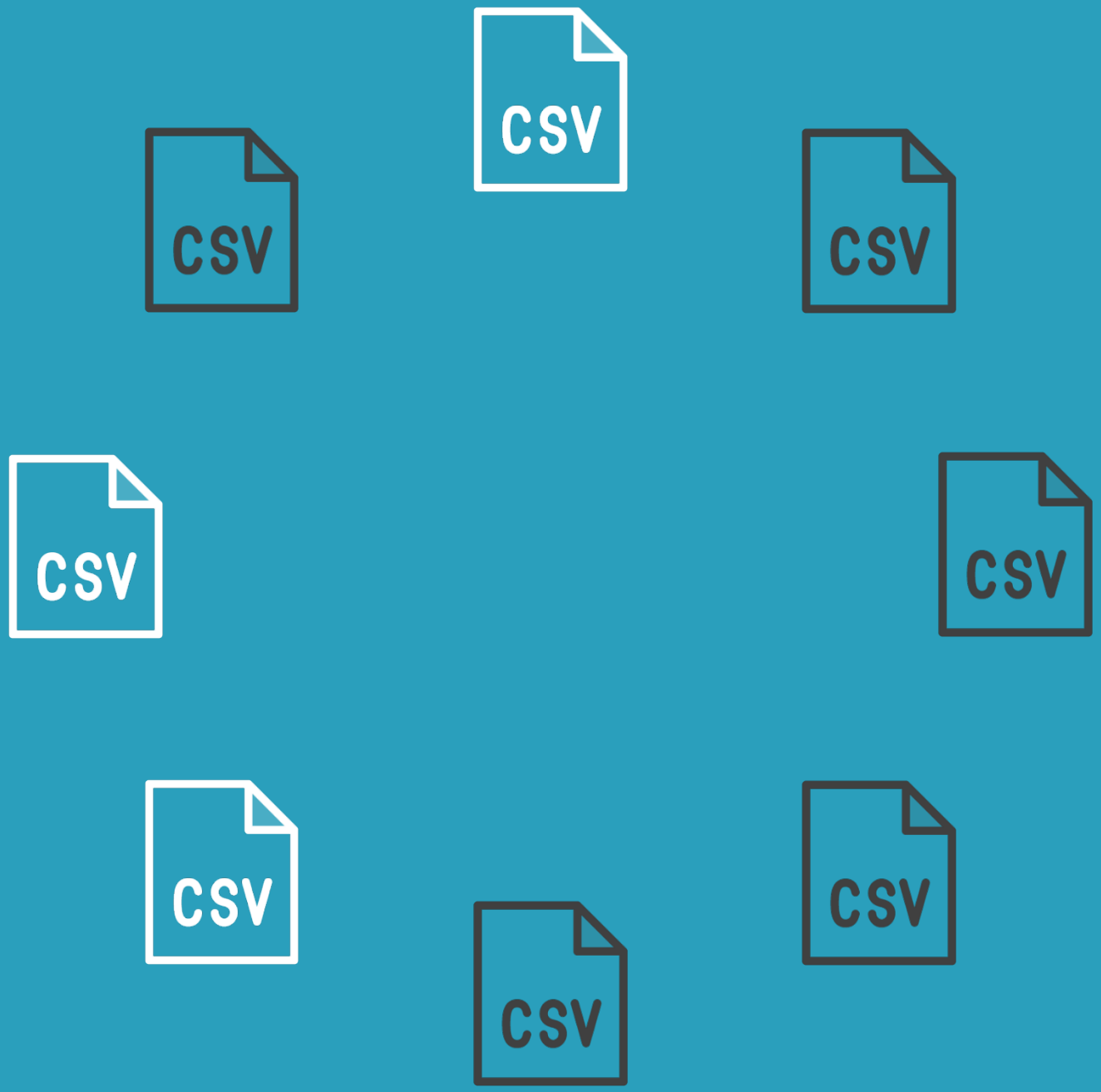
For more information about the Hockey Databank project please visit:
https://groups.yahoo.com/groups/hockey-databank

NHL

National Hockey League




CSV
CSV
CSV
CSV
CSV
CSV
CSV
CSV


CSV
CSV
CSV
CSV
CSV
CSV
CSV
CSV

README


INFO


Players

Teams

Player Data

Master.csv

playerID	coachID	firstName	lastName	birthYear	birthMon	birthDay
adamru01		Russ	Adam	1961	5	5
bowneri01	bowneri01	Rick	Bowness	1955	1	25
	lapoiro01c	Ron	Lapointe	1949	11	12

Scoring.csv

playerID	year	tmID	pos	GP	G	A	Pts	SOG
amontto01	2006	CAL	R	81	10	20	30	139
amontto01	2005	CAL	R	80	14	28	42	155
terrigr01	1980	LAK	L	73	12	25	37	92
harkibr01	1994	BOS	C	1	0	1	1	1

Scoring.csv

Year

playerID	year	tmID	pos	GP	G	A	Pts	SOG
amontto01	2006	CAL	R	81	10	20	30	139
amontto01	2005	CAL	R	80	14	28	42	155
terrigr01	1980	LAK	L	73	12	25	37	92
harkibr01	1994	BOS	C	1	0	1	1	1

Scoring.csv

ID of the team

playerID	year	tmID	pos	GP	G	A	Pts	SOG
amontto01	2006	CAL	R	81	10	20	30	139
amontto01	2005	CAL	R	80	14	28	42	155
terrigr01	1980	LAK	L	73	12	25	37	92
harkibr01	1994	BOS	C	1	0	1	1	1

Scoring.csv

Position

playerID	year	tmID	pos	GP	G	A	Pts	SOG
amontto01	2006	CAL	R	81	10	20	30	139
amontto01	2005	CAL	R	80	14	28	42	155
terrigr01	1980	LAK	L	73	12	25	37	92
harkibr01	1994	BOS	C	1	0	1	1	1

Scoring.csv

Games played

playerID	year	tmID	pos	GP	G	A	Pts	SOG
amontto01	2006	CAL	R	81	10	20	30	139
amontto01	2005	CAL	R	80	14	28	42	155
terrigr01	1980	LAK	L	73	12	25	37	92
harkibr01	1994	BOS	C	1	0	1	1	1

Scoring.csv

Goals

playerID	year	tmID	pos	GP	G	A	Pts	SOG
amontto01	2006	CAL	R	81	10	20	30	139
amontto01	2005	CAL	R	80	14	28	42	155
terrigr01	1980	LAK	L	73	12	25	37	92
harkibr01	1994	BOS	C	1	0	1	1	1

Scoring.csv

Assists

playerID	year	tmID	pos	GP	G	A	Pts	SOG
amontto01	2006	CAL	R	81	10	20	30	139
amontto01	2005	CAL	R	80	14	28	42	155
terrigr01	1980	LAK	L	73	12	25	37	92
harkibr01	1994	BOS	C	1	0	1	1	1

Scoring.csv

Points = Goals + Assists

playerID	year	tmID	pos	GP	G	A	Pts	SOG
amontto01	2006	CAL	R	81	10	20	30	139
amontto01	2005	CAL	R	80	14	28	42	155
terrigr01	1980	LAK	L	73	12	25	37	92
harkibr01	1994	BOS	C	1	0	1	1	1

Scoring.csv

Shots on goal

playerID	year	tmID	pos	GP	G	A	Pts	SOG
amontto01	2006	CAL	R	81	10	20	30	139
amontto01	2005	CAL	R	80	14	28	42	155
terrigr01	1980	LAK	L	73	12	25	37	92
harkibr01	1994	BOS	C	1	0	1	1	1

Team Data

Teams.csv

year	tmID	name
1999	PHO	Phoenix Coyotes
2010	AND	Anaheim Ducks
2011	NJD	New Jersey Devils

TeamSplits.csv

year	tmID	OctW	OctL	OctT	OctOL
2008	DAL	4	5		2
2008	DET	7	2		2
2008	EDM	4	4		1

TeamSplits.csv

October wins

year	tmID	OctW	OctL	OctT	OctOL
2008	DAL	4	5		2
2008	DET	7	2		2
2008	EDM	4	4		1

TeamSplits.csv

October losses

year	tmID	OctW	OctL	OctT	OctOL
2008	DAL	4	5		2
2008	DET	7	2		2
2008	EDM	4	4		1

TeamSplits.csv

October ties

year	tmID	OctW	OctL	OctT	OctOL
2008	DAL	4	5		2
2008	DET	7	2		2
2008	EDM	4	4		1

TeamSplits.csv

October overtime losses

year	tmID	OctW	OctL	OctT	OctOL
2008	DAL	4	5		2
2008	DET	7	2		2
2008	EDM	4	4		1

Year 2010

September 2010 – April 2011